University of Alberta

Assembly and Structural Analysis of the Mammalian pre-Spliceosome Commitment Complex

by

Oliver Andrew Kent ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Biochemistry

Edmonton, Alberta
Fall, 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-95952-X
Our file Notre référence
ISBN: 0-612-95952-X

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Dedicated to my Wife Samantha...

...eventually all the pieces come together.

Acknowledgement:

I would like to acknowledge many people who helped me, guided me, and affected the outcome of this thesis in one or another.

My Wife Samantha has given me her 100% support during my studies. I thank her for being there for me and especially being "here" for me. Samantha was very selfless and understanding when she "eagerly" agreed to move to Edmonton from Toronto. I can never fully express my gratitude and appreciation to her, but having Samantha in Edmonton with me over the past three years has made my life in and out of the lab complete.

Dr. Andrew MacMillan has given me his unconditional support and guidance during graduate school. Andrew has made my experiences in the lab only positive. I was very fortunate to have crossed paths with Andrew (thanks then are needed to Dr. Rob Batey for not having room in his fume hood!). Over the past five plus years my relationship with Andrew has developed from supervisor-student into one of friend-friend. I highly respect Andrew and believe his supervision is second to none.

My parents, Pixie and Alan Roney, and Colin Kent have given me both support and encouragement during my time in graduate school. Without the full support (financial) of Pixie and Alan I would have not had the opportunity to go to University in the beginning; I am extremely grateful that they gave me the chance. Although Mom laughed when I told her I wanted to become a doctor, I have now proven that I am very capable. I would also like to express thanks to my Dad Colin, who has been very interested in my career and continues to take interest in my publications and research, even though he doesn't really understand what it is I am researching! Annabelle, Ashley, and Samantha (the sibs) have given me their encouragement and advice on both personal and academic matters for which I am very grateful. I would also like to thank my mother-in-law Barb Smith for many encouraging words and being on my side when needed (you know what I mean!).

I would like to acknowledge the professors who participated in my defense. Thanks to both Dr. Mark Glover and Dr. Kevin Wilson for being members of my supervisory committee. Collaborative group meetings between the MacMillan, Glover, and Wilson labs proved to very rewarding in getting my project moving in a forward direction. I am regretful that Mark was unable to attend my thesis defense, but appreciate all his help and suggestions that led to it. I was very lucky to have met Kevin Wilson, who coincidently had a lot of experience with Fe-BABE. I thank Kevin for sharing his Fe-BABE knowledge with me and providing me RT conditions that work flawlessly. Kevin was a valuable colleague during analysis of our early data.

Thanks to Dr. Ben Blencowe (University of Toronto) for coming to Edmonton and being an excellent external examiner. Ben provided a thorough and extremely well documented review of my thesis that was a lot of help during the revisions stage. I would like to thank Dr. Mike Schultz for substituting for Dr. Glover as a member of my committee, and Dr. Laura Frost for being a member of my defense committee.

I would like to thank all members of the MacMillan lab, both past and present (even those who didn't acknowledge me...D.J,L.F....whatever!). Dr. Steven Chaulk, a previous MacMillan lab member, has helped and guided me immensely during my first couple of years in graduate school. Steven is not only a good colleague but also a good friend. The current MacMillan lab members (also good friends) Kaari Chilibeck (face it your calculator is mine and I really won the hockey pool!), Matt Schellenberg (the almost new Dad who scoops D.R. projects!), Dustin Ritchie (Dr. Fu who is the self proclaimed coolest person…ever!), Chelsea Topping (you can't have my bench just yet!), and Tao Wu (the new guy), thanks for your comments and suggestions to my research, for proof reading my thesis, and attending practice talks.

I would like to thank some of the other members of the department of biochemistry that have contributed to my research or studies. Thanks to Dave Arthur, Ross Edwards, Alex Ghetu, Ruth Green, Ryan Hoffman, Jason Lamoureux, Brian Mark, Scott Williams, and anyone else I may have non-intentionally forgot. Thanks to both Marion Benedict and Susan Smith for their incredible secretarial services. These two ladies made the organization of dates, people, and rooms a smooth process.

Finally, I would also like to thank all my non-university associated friends from both Toronto and Edmonton for all the support and encouragement they have given me. Thanks to Shaun and Sue Ryan, Chris and Cheryl Williams, and Mike and Laura Losse for being there for me while I am so far from home. I would like to thank Joan Noddings her wonderful sister Sue, her brother Greg, and the bratty kids Matt and Dane (alright hyper but maybe not bratty?) for treating Samantha and I like part of their family. Thanks to Mike and Joanne Allen, thanks to Dennis (toad) and everyone at Charles Williams, and thank you to anyone I may have forgotten. Everyone on this list has been there for me in one way or another during my time away from the lab (which isn't much) and provided me with a much needed divergence from science.

Table of Contents:

- Introduction. Pre-mRNA splicing and spliceosome assembly. 1
  - I-1. Introduction to pre-mRNA splicing. 2
    - I-1.1. Pre-mRNA splicing. 2
    - I-1.2. Pre-mRNA intronic sequence elements. 3
    - I-1.3. Spliceosome assembly and formation of the active site. 5
    - I-1.4. Recognition of the 5' and 3' splice sites. 9
    - I-1.5. Formation of the pre-spliceosome commitment complex. 14
    - I-1.6. The commitment complex component U2AF. 14
  - I-2. Thesis overview and future directions. 17
    - I-2.1. Organization of the pre-mRNA substrate in the commitment complex. 17
  - I-3. References. 18

- Chapter 1. Early organization of pre-mRNA during spliceosome assembly. 26
  - 1-1. Introduction. 27
    - 1-1.1. Spliceosome assembly. 27
    - 1-1.2. Intron structure within pre-spliceosomes. 28
  - 1-2. Results and Discussion. 29
    - 1-2.1. Synthesis and derivatization of Fe-BABE modified pre-mRNA. 29
    - 1-2.2. Mapping pre-mRNA structure in the pre-spliceosomal E complex. 35
    - 1-2.3. Factor and sequence requirements in pre-spliceosome assembly. 36
    - 1-2.4. Implication for spliceosome assembly. 39
  - 1-3. Materials and Methods. 41
    - 1-3.1. Synthesis of Fe-BABE pre-mRNA substrates. 41
    - 1-3.2. Pre-mRNA splicing and complex formation. 45
    - 1-3.3. Probing pre-mRNA structure in splicing complexes. 47
  - 1-4. References. 48

- Chapter 2. Structuring of the 3' splice site by U2AF65. 53
  - 2-1. Introduction. 54
    - 2-1.1. Organization at the 3' splice site. 54
    - 2-1.2. The splicing factor U2AF. 55
  - 2-2. Results and Discussion. 57
    - 2-2.1. Functionalization of U2AF65 with the chemical nuclease Fe-EDTA. 57
    - 2-2.2. Protection of the branch by the RS domain of U2AF65. 59
    - 2-2.3. Mapping U2AF65•RNA interaction using tethered Fe-EDTA. 61
    - 2-2.4. Structure of higher order complexes at the 3' splice site. 63
    - 2-2.5. Model of RNA recognition by polypyrimidine binding factors. 64
  - 2-3. Materials and Methods. 67
    - 2-3.1. Preparation of U2AF binding substrates. 67
    - 2-3.2. Expression and derivatization of U2AF65-based probes. 68
    - 2-3.3. Binding and affinity cleavage. 72
  - 2-4. References. 73

- Chapter 3. Characterization of a U2AF independent commitment complex (E'). 78
  - 3-1. Introduction. 79
    - 3-1.1. The commitment complex. 79
  - 3-2. Results and Discussion. 81
    - 3-2.1. Identification of a novel pre-spliceosomal complex. 81
    - 3-2.2. Relationship of E' to the pre-spliceosome. 83
    - 3-2.3. Directed hydroxyl radical probing of pre-mRNA structure in E'. 85
    - 3-2.4. Detection of branch binding protein in E' complex. 87
    - 3-2.5. The E' complex is the mammalian equivalent to the yeast CC1. 89
    - 3-2.6. Structural organization of the E' complex. 91
  - 3-3. Materials and Methods. 92
    - 3-3.1. Synthesis of pre-mRNA substrates. 92
    - 3-3.2. Analysis of pre-spliceosome complexes and pre-mRNA splicing. 94
    - 3-3.3. Affinity cleavage and UV crosslinking. 96
  - 3-4. References. 97

Chapter 4. Commitment complex organization in the absence of U1 snRNA•pre-mRNA base pairing. 102
- 4-1. Introduction. 103
  - 4-1.1. Association of U1 snRNP with the 5' splice site. 103
- 4-2. Results and Discussion. 105
  - 4-2.1. U1 snRNP is a commitment complex component. 105
  - 4-2.2. U1 snRNP association with the 5' splice site. 107
  - 4-2.3. Splicing and complex analysis in truncated U1 snRNA extracts. 108
  - 4-2.4. Two conformations at the 5' splice site. 112
  - 4-2.5. U1-C interaction precedes U1 snRNA base pairing. 114
- 4-3. Materials and Methods. 117
  - 4-3.1. Synthesis of pre-mRNA substrates. 117
  - 4-3.2. Analysis of pre-mRNA splicing and complex formation. 118
  - 4-3.3. Analysis of Fe-BABE cleavage reactions. 120
- 4-4. References. 121

Chapter 5. Future directions: transcription and RNA processing. 125
- 5-1. Future directions. 126
  - 5-1.1. Applications to spliceosome assembly. 126
  - 5-1.2. Transcription and pre-mRNA splicing are coordinated events. 127
  - 5-1.3. Pre-mRNA splicing and A to I RNA editing. 128
- 5-2. References. 129

Appendix *I*. Kinetic analysis of the M1 RNA folding pathway. 133
- *I*-1. Introduction. 134
  - *I*-1.1. RNase P M1 RNA structure and folding. 134
- *I*-2. Results and Discussion. 137
  - *I*-2.1. Endoribonuclease activity and folding M1 RNA. 137
  - *I*-2.2. Kinetic footprinting of M1 RNA with peroxynitrous acid. 140
  - *I*-2.3. Implications of kinetic analysis of M1 RNA folding and structure. 143
- *I*-3. Materials and Methods. 147
  - *I*-3.1. Synthesis of M1 RNA and ptRNA substrates. 147
  - *I*-3.2. M1 RNA activity and folding. 149
  - *I*-3.3. Kinetic footprinting M1 RNA. 150
- *I*-4. References. 151

Appendix *II*. Synthesis of a convertible cytosine ribonucleotide. 154
- *II*-1. Introduction. 155
  - *II*-1.1. Modified nucleosides. 155
  - *II*-1.2. The convertible nucleotide strategy. 156
- *II*-2. Results and Discussion. 158
  - *II*-2.1. Synthesis of 4-chlorophenyl uridine. 158
  - *II*-2.2. Synthesis of a functionalized RNA. 160
- *II*-3. Materials and Methods. 162
  - *II*-3.1. Synthesis of 4-O-(4-chlorophenyl) uridine phosphoramidite. 162
  - *II*-3.2. Synthesis of a functionalized RNA. 166
- *II*-4. References. 168

List of Tables:

Appendix *I*.

Table *I*-1. Rates of folding for regions of M1 RNA observed using peroxynitrous acid as a kinetic footprinting reagent. 144

List of Figures:

Introduction.

I-1. The pre-mRNA splicing reaction. 4
I-2. The spliceosome assembly cycle. 6
I-3. RNA-RNA rearrangements during spliceosome assembly. 8
I-4. Organization of U1 snRNP components. 11
I-5. Organization of commitment complex components. 13
I-6. The splicing factor U2AF. 15

Chapter 1.

1-1. Synthesis of a branch oligonucleotide site-specifically modified with Fe-BABE. 31
1-2. Synthesis and splicing of pre-mRNA site-specifically modified with Fe-BABE. 32
1-3. Directed hydroxyl radical probing in the spliceosome and pre-spliceosome. 34
1-4. Factor and sequence dependence of E complex formation and Fe-BABE mediated cleavages. 37
1-5. Early organization of pre-mRNA in spliceosome assembly. 40

Chapter 2.

2-1. U2AF organization at the 3' splice site. 56
2-2. Modification of U2AF65 based probes with Fe-EDTA. 58
2-3. Mobility shift and RNase mapping of U2AF65•RNA interaction. 60
2-4. Directed hydroxyl radical cleavage of RNA by U2AF65-Fe-EDTA. 62
2-5. U2AF65 bends the polypyrimidine tract. 65

Chapter 3.

3-1. Identification of a novel pre-spliceosome complex (E') that contains U1 snRNP. 82
3-2. Commitment of pre-mRNA to splicing in the absence of U2AF. 84
3-3. Directed hydroxyl radical cleavage of ATP independent pre-spliceosome complexes. 86
3-4. Site-specific protein crosslinking at the branch region. 88
3-5. E complex assembly. 90

Chapter 4.

4-1. Analysis of U1 snRNP depleted and U1 snRNP blocked nuclear extracts. 106
4-2. RNase H digestion of U1 snRNA. 109
4-3. Sequestering of pre-mRNA to pre-spliceosome complexes in the absence of U1 snRNA-pre-mRNA base pairing. 110
4-4. Directed hydroxyl radical cleavage in Δ1-12-U1 snRNA pre-spliceosome complexes. 113
4-5. Two conformations at the 5' splice site. 115

Appendix *I*.

I-1. RNase P. 136
I-2. M1 RNA enzymatic activity assay. 138
I-3. Endoribonuclease assay of M1 folding. 139
I-4. Magnesium dependent protection of M1 RNA. 141
I-5. Kinetic footprinting of the M1 RNA folding pathway using peroxynitrous acid. 142
I-6. Folding pathway of RNase P M1 RNA. 145

Appendix *II*.

II-1. The convertible nucleotide strategy. 157
Scheme *II*-1 159
Scheme *II*-2 159
II-2. Deprotection of benzoyl protected cytosine. 161

List of Abbreviations:

Δ	delta, depleted
A	adenosine
AG	3' splice site dinucleotide
ATP	adenosine triphosphate
BBP	branch binding protein
BPS	branch point sequence
BSA	bovine serum albumin
C	cytosine
CC	commitment complex
CIAP	calf intestinal alkaline phosphatase
ClφU	4-chlorophenyl uridine
CP	creatine phosphate
ddNTP	dideoxy nucleoside triphosphate
dNTP	deoxy nucleoside triphosphate
DTT	dithiolthreitol
E complex	early complex
EDTA	ethylene diamine tetraacetic acid
Fe-BABE	bromoacetamidobenzyl-EDTA
Fe-EDTA	iron EDTA
G	guanidine
H complex	hnRNP complex
His$_6$	polyhistidine tag
hnRNP	heterogeneous nuclear ribonucleoprotein
IVS	intervening sequence
KO	knock out
MPA	mercaptopropionic acid
mRNA	messenger RNA
N	any nucleotide
NE	nuclear extract
NTP	nucleoside triphosphate
OMe	methoxy
PAC	phenoxyacetyl
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PIP	precursor in pieces
PPT	polypyrimidine tract
Pre-mRNA	Precursor mRNA
R	purine
RRM	RNA recognition motif
RS	arginine-serine rich domain
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
SELEX	systematic evolution of ligands by exponential enrichment
SF1	splicing factor 1
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
sxl	sex lethal
TBE	tris borate EDTA
tRNA	transfer RNA
U	uridine
U2AF	U2 snRNP auxiliary factor
WT	wild type
Y	pyrimidine

Introduction.

Pre-mRNA splicing and spliceosome assembly.

Introduction. Pre-mRNA splicing and spliceosome assembly.

I-1. Introduction.

I-1.1. Pre-mRNA splicing.

In 1977, work by the Sharp and Roberts labs revealed that genes of higher organisms are "split" or present in several distinct segments along the DNA molecule (Berget et al., 1977; Chow et al., 1977). The coding regions of the gene are separated by non-coding DNA that is not involved in protein expression. The split gene structure was found when adenoviral mRNAs were hybridized to endonuclease cleavage fragments of single stranded viral DNA (Berget et al., 1977). It was observed that the mRNAs of the mRNA-DNA hybrids contained 5' and 3' tails of non-hydrogen bonded regions. When larger fragments of viral DNAs were used, forked structures of looped out DNA were observed when hybridized to the viral mRNAs. It was realized that the looped out regions or intervening sequences are excised from the precursor mRNAs in a process Sharp coined "splicing". The split gene structure was subsequently found to be common to most eukaryotic genes.

It is now known that the removal of intervening sequences from pre-mRNAs is catalyzed by the spliceosome: a 60S biochemical machine that consists of both protein and RNA components (Krämer, 1996). The spliceosome directs the removal of intron sequences from the pre-mRNA and subsequent ligation of the exons to produce a functional mRNA (Burge et al. 1999; Staley and Guthrie, 1998). Pre-mRNA splicing takes place through two sequential transesterification reactions (Figure I-1). In the first step, the 2' OH of a conserved adenosine within the branch point region performs a nucleophilic displacement at the 5' splice site to produce the lariat intermediate and the liberated 5' exon. The 5' end of the intron is joined via a 2'-5' phosphodiester linkage to the adenosine residue within the branch point region creating the lariat intermediate. In the second step, the 3' OH of the 5' exon performs a nucleophilic displacement at the 3' splice site to produce the lariat product and the ligated exons. It has been determined using chiral phosphorothioate substitutions at the 5' and 3' splice site that both steps of splicing are inhibited by the R_p phosphorothiate diastereomer but not by the S_p (Moore and Sharp, 1993). In addition, both chemical steps occur as single inline S_N2 nucleophilic displacement reactions, analogous to the mechanism of group I self-splicing introns (McSwiggen and Cech, 1989; Rajagopal et al., 1989). Based on these observations it has been suggested that the splicing machinery probably shifts between two active sites in the catalysis of the two transesterification reactions (Moore and Sharp, 1993), and is consistent with the idea that pre-mRNA splicing is RNA catalyzed (Moore et al., 1993).

Both chemical steps of splicing are catalyzed by the spliceosome, which consists of the U1, U2, and U4/U5/U6 snRNPs (small nuclear ribonucleoprotein particles) each containing a unique snRNA and associated proteins bound to the pre-mRNA substrate (Burge et al., 1999; Krämer, 1996; Staley and Guthrie, 1998). Assembly of the spliceosome active site on the pre-mRNA proceeds through several stages and is directed by conserved sequences at the splice sites and within the intron as well as other sequences in the pre-mRNA.

I-1.2. Pre-mRNA intronic sequence elements.

The exact sites for the splicing transesterifications of mammalian pre-mRNAs are defined by consensus sequence elements at the 5' and 3' splice sites. The 5' splice site consensus is relatively short: AG/$\underline{\text{GU}}$RAGU (where R= purine, Y=pyrimidine, N=any nucleotide, and / indicates a splice site). In contrast, the 3' splice site is defined by three separate sequence elements: the branch point sequence, the polypyrimidine tract, and the 3' splice site consensus. In mammals, the highly degenerate branch sequence is YNYUR$\underline{\text{A}}$C (where $\underline{\text{A}}$ is the branch point adenosine), and the 3' splice site consensus is Y$\underline{\text{AG}}$/G (Moore et al., 1993). The polypyrimidine tract is found between the branch point sequence and the 3' splice site and is typically 14-40 nucleotides in length. Typical polypyrimidine tracts function to increase the efficiency of branch point utilization and 3' splice site recognition. However, these two functions can be separated based on the location of pyrimidine stretches within a given tract (Reed, 1989).

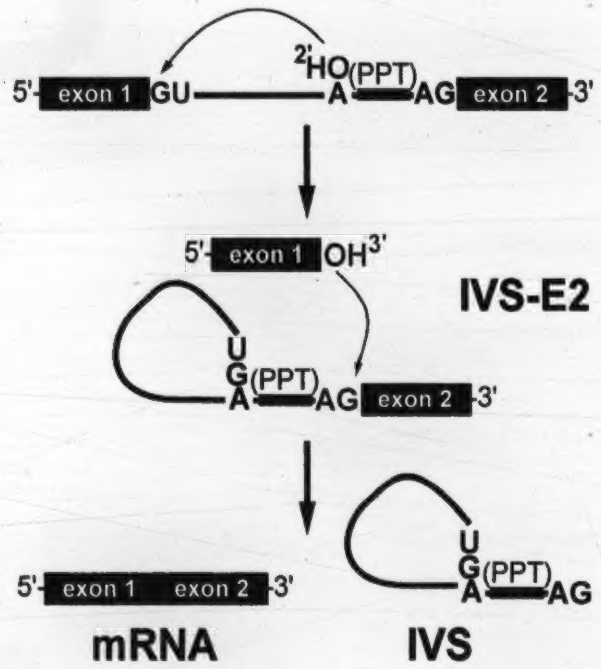


Figure I-1. The pre-mRNA splicing reaction. Pre-mRNA splicing proceeds through two sequential transesterification reactions. The mammalian pre-mRNA substrate contains a conserved GU at the 5' splice site, YAG at the 3' splice site, and a branch point adenosine (A) within the branch region followed by the polypyrimidine tract (PPT). The splicing products are the lariat intermediate (IVS-E2), the lariat product (IVS) and the mRNA.

Interestingly, introns can be classified on the basis of long and short polypyrimidine tracts: AG independent and AG dependent respectively (Moore, 2000; Reed, 1989). In AG independent introns, the formation of the spliceosome and first catalytic step of splicing do not depend on the conserved AG dinucleotide at the 3' splice site. In contrast, AG dependent introns do not form spliceosome complexes or undergo either step of splicing if the AG at the 3' splice site is mutated (Reed, 1989).

I-1.3. Spliceosome assembly and formation of the active site.

The canonical model for formation of the spliceosome active site involves an ordered, stepwise assembly of discrete snRNP particles on the pre-mRNA substrate (Figure I-2). The first recognition of pre-mRNAs involves U1 snRNP binding to the 5' splice site of the pre-mRNA and other non-snRNP associated factors to form the commitment complex or early (E) complex in mammals (Jamison et al., 1992; Seraphin and Rosbash, 1989). The commitment complex is an ATP independent complex that commits the pre-mRNA to the splicing pathway (Legrain et al., 1988). U2 snRNP is recruited to the branch region through interactions with the E complex component U2AF (U2 snRNP auxiliary factor) and possibly U1 snRNP. In an ATP dependent reaction, U2 snRNP becomes tightly associated with the branch point sequence (BPS) to form complex A. A duplex formed between U2 snRNA and the pre-mRNA branch region bulges out the branch adenosine specifying it as the nucleophile for the first transesterification (Query et al., 1994). The presence of a pseudouridine residue in U2 snRNA nearly opposite the branch site results in an altered conformation of the RNA-RNA duplex upon U2 snRNP binding. Specifically, the altered structure of the duplex induced by the pseudouridine places the 2' OH of the bulged adenosine in a favorable position for the first step of splicing (Newby and Greenbaum, 2002). The U4/U5/U6 tri-snRNP is recruited to the assembling spliceosome to form complex B, and following several rearrangements, complex C (the spliceosome) is activated for catalysis (Burge et al., 1999; Staley and Guthrie, 1998). It is unclear how the triple snRNP is recruited to complex A but this process may be mediated through protein-protein interactions and/or base pairing interactions between U2 snRNA and U6 snRNA.

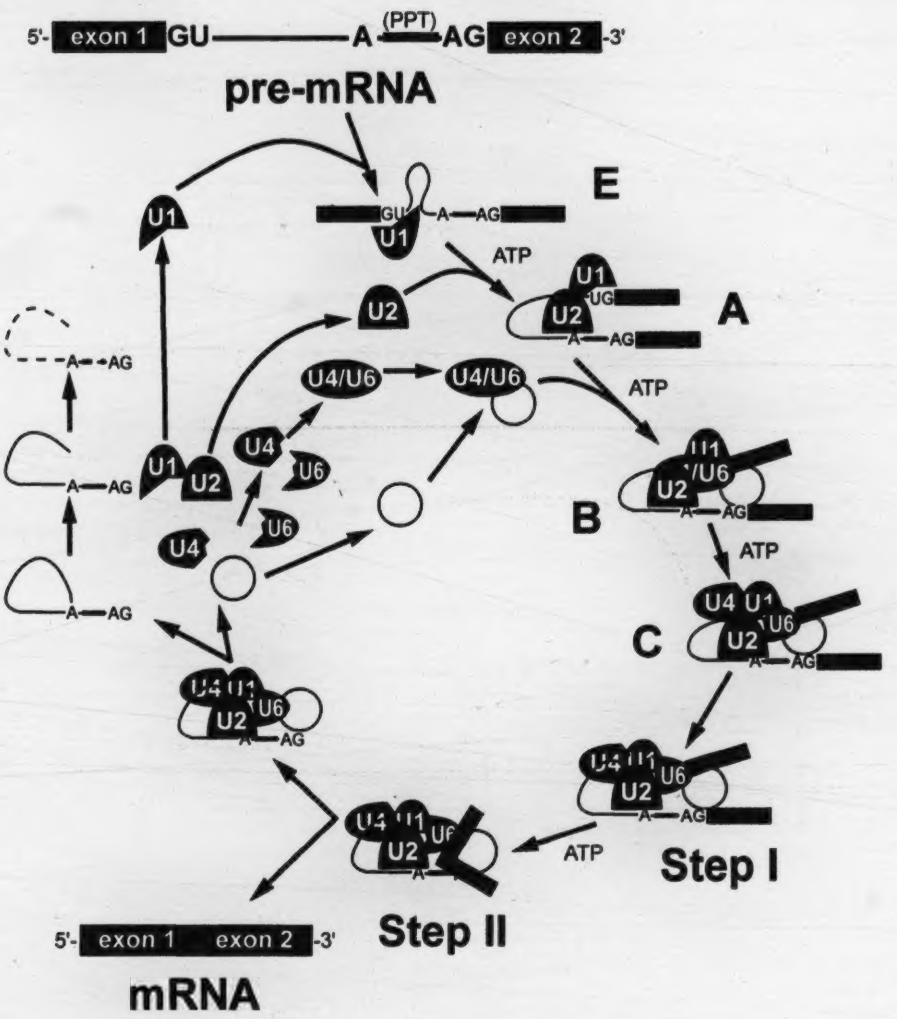


Figure I-2. The spliceosome assembly cycle. The assembly of the mammalian spliceosome proceeds through the ATP independent early complex (E), followed an ATP dependent formation of the A, B, and C complexes. Chemical steps I and II occur after rearrangements within the C complex (Adapted from Moore et al., 1993).

The U5 snRNP interacts with sequences at the 5' and 3' splice sites via the invariant loop of U5 snRNA (Newman et al., 1995) and U5 protein components interact with the 3' splice site region (Chiara et al., 1997).

Upon recruitment of the triple snRNP several RNA-RNA rearrangements precede the first catalytic step and further rearrangements occur in the catalytically active spliceosome. Several of the RNA-RNA interactions are mutually exclusive (Figure I-3); however, it is not known what triggers these interactions or the order of these rearrangements. The first rearrangement is probably the displacement of U1 snRNP from the 5' splice site and formation of a U6 snRNA interaction (Figure I-3A). It is known that U1 snRNP is only weakly associated with fully formed spliceosomes (Moore et al., 1993), and U1 snRNP is inhibitory to the formation of a U6-5' splice site interaction on a model substrate oligonucleotide containing a short 5' exon and 5' splice site (Konforti et al., 1993). Binding of U2 snRNP to the branch point sequence (BPS) is one example of an RNA-RNA interaction displacing a protein-RNA interaction. Upon recruitment of U2 snRNP, the branch binding protein SF1 in the commitment complex is displaced since the binding site of U2 snRNA and SF1 are mutually exclusive events (Figure I-3B). Within the U2 snRNA there are other mutually exclusive rearrangements that occur between competing conformations. For example, in the active form, stem loop IIa is favored (Figure I-3C); in the inactive form a mutually exclusive interaction between the loop and a downstream sequence predominates (Staley and Guthrie, 1998). It is unclear how U4 is displaced from U6 snRNA; a number of RNA helicases have been implicated in spliceosome assembly and may function to unwind U4/U6 and promote the formation of a U2/U6 snRNA interaction (Figures I-3D and I-3E). The interactions of U4/U6 stem loops I and II dissociate and the freed stem loop II region of U6 folds on itself to form an intramolecular stem loop (Figure I-3D) and U4 is no longer required in further spliceosome assembly. The freed stem loop I region of U6 base pairs with U2 snRNA forming the U2/U6 helix I (Figure I-3E). However, the helix I structure is mutually exclusive with the 3' half of an internal 5' stem loop region of U2 snRNA (Figure I-3E).

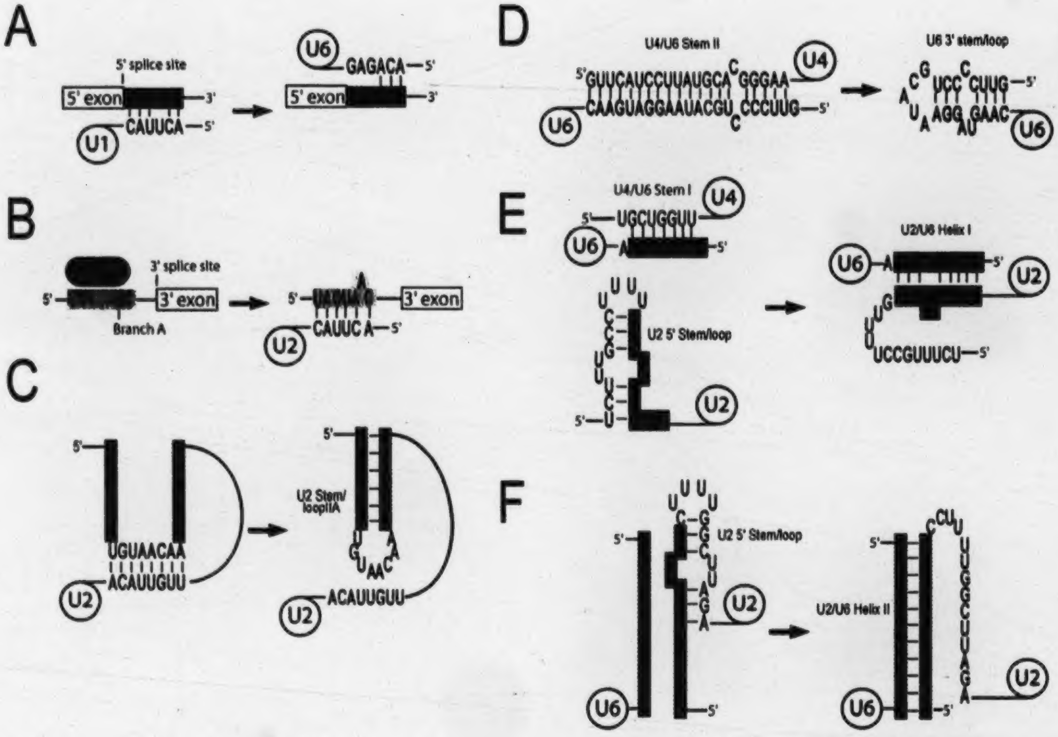


Figure I-3. RNA-RNA rearrangements within the spliceosome. Formation of the active spliceosome proceeds through many mutually exclusive interactions. **(A)** The exchange of U1 for U6 at the 5' splice site. **(B)** The exchange of branch binding protein (BBP) for U2 snRNA. **(C)** An intramolecular U2 rearrangement. **(D)** Disruption of U4/U6 stem II and formation of U6 3' stem/loop. **(E)** Disruption of U4/U6 stem I and formation of U2/U6 helix I. **(F)** Disruption of the U2 5' stem/loop and formation of U2/U6 helix II. Sequences are from yeast snRNAs. (Adapted from Staley and Guthrie, 1998).

The release of this helix resulting in formation of U2/U6 helix I also frees the 5' half of the internal 5' step loop of U2 snRNA enabling the formation of U2/U6 helix II (Figure I-3F). Although these base pairing interactions have been elucidated, the timing of these rearrangements with respect to each other is yet to be determined.

Using a 2'-caged adenosine which blocks the first step of splicing until the caged pre-mRNA has been photolyzed has demonstrated that once the spliceosome has assembled, catalytic step I occurs in the absence of ATP (Chaulk and MacMillan, 2000). This result has effectively separated spliceosome assembly from catalysis. Thus, assembly of the spliceosome active site poises the 2'-OH of the branch adenosine in position to carry out the first transesterification reaction without further rearrangement. However, following photolysis of the caging group and release of the 2'-OH of the branch point adenosine, the spliceosome catalyzed step I but was unable to catalyze step II of splicing in the absence of ATP. Therefore, following the first catalytic step of splicing, ATP dependent rearrangements must occur to facilitate the second step of splicing.

RNA-RNA interactions have been implicated in configuring the reactants for the second catalytic step: the highly conserved loop of U5 snRNA contacts the 5' and 3' ends of the intron (Newman, 1997; Newman et al., 1995) and has been suggested to align the exons for the second transesterification reaction. Furthermore, the protein factor Slu7 has been shown to be essential for correct 3' splice site choice by interacting with the freed 5' exon and aligning it for correct attack at the 3' splice site (Chau and Reed, 1999).

I-1.4. Recognition of the 5' and 3' splice sites.

Formation of the pre-spliceosome and recognition of the pre-mRNA substrate occurs in an ATP independent fashion and involves a minimal set of factors that associate with the 5' splice site and polypyrimidine tract/3' splice site within the pre-mRNA substrate. Early recognition of the pre-mRNA substrate involves the association of U1 snRNP with the 5' splice site to form the commitment complex (CC, yeast) or early complex (E, mammals). Several ATP independent complexes have previously been identified to form in HeLa nuclear extract (Michaud and Reed, 1993). These include the hnRNP complex (H) which is formed by the non-specific association of ribonucleoproteins with RNA, and a "–ATP" complex that has been identified to contain U2 snRNP (Jamison and Garcia-Blanco, 1992).

U1 snRNP has been implicated in the initiation of spliceosome assembly through base pairs between the single stranded 5' arm of U1 snRNA and the conserved stretch of six nucleotides at the 5' splice site (Moore et al., 1993). Mammalian U1 snRNP consists of ten different proteins and a 165 nucleotide RNA molecule (Figure I-4). U1 snRNP contains 7 Sm proteins that are proposed to form a ring structure around the Sm binding site of the U1 snRNA, and three U1 specific proteins U1-70K, U1-A, and U1-C (Figure I-4A). Recently, the cryomicroscopy structure of the U1 snRNP particle has been determined allowing a structural model that describes the three dimensional arrangement of proteins and RNA in U1 snRNP (Figure I-4B; Stark et al., 2001). The structure consists of a ring-shaped main body containing the Sm protein core, and two large protuberances corresponding to U1-70K and U1-A proteins (Figure I-4B).

Several groups have demonstrated that the initial recognition of the 5' splice site is mediated by the association of U1-C with a conserved sequence at the 5' splice site (Du and Rosbash, 2002; Heinrichs et al., 1990; Rossi et al., 1996). This initial association of U1 snRNP with the 5' splice site is then followed by a stable base pairing interaction of the U1 snRNA with the conserved sequence at the 5' splice site (Nilsen, 1998). It is believed that U1-C may potentiate the base pairing interaction of U1 snRNA with the pre-mRNA (Tang et al., 1997). Indeed, the direct contact of U1-C with the 5' splice site is rapid and is almost irreversible and digestion of the 5' arm of U1 snRNA with RNase H has no significant effect on the association of U1 snRNP proteins with the 5' splice site (Rossi et al., 1996). These results provide evidence that U1 snRNP specific proteins are required for 5' splice site interaction but that U1 snRNA base pairing stabilizes this interaction. In addition, U1 snRNP may be involved in recruiting U2AF to the polypyrimidine tract through indirect interactions with the SR protein SC35.

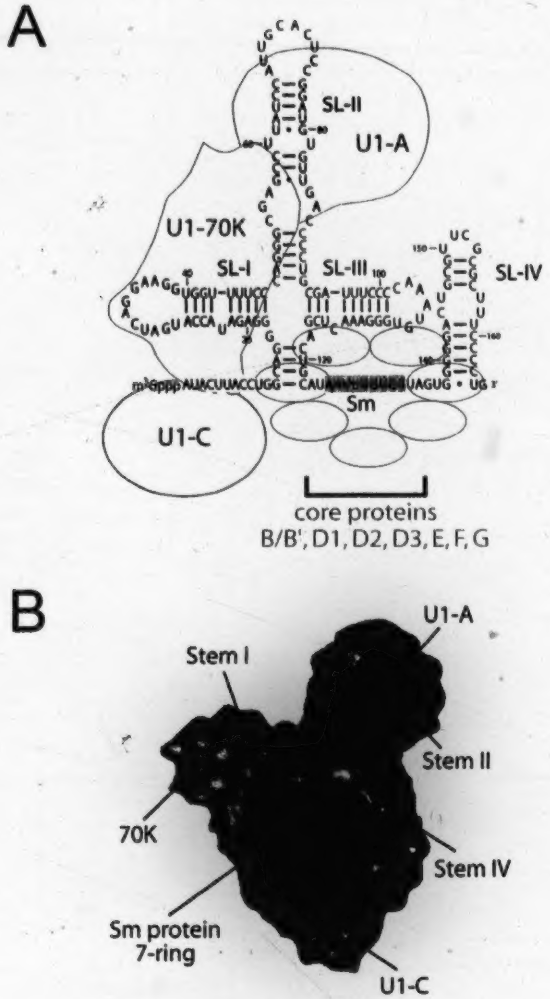


Figure I-4. Organization of U1 snRNP components. **(A)** Schematic organization of U1 snRNP components. The human U1 snRNP contains a 165 nucleotide snRNA with a four stem loop structure. The U1 snRNA 5' splice site recognition sequence is shown (yellow). U1 snRNP specific protein U1-70K interacts with stem loop I, U1-A interacts with stem loop II, and the seven core proteins form a ring structure around the Sm binding site (brown). U1-C is loosely associated with U1 snRNP. **(B)** The electron cryomicroscopy structure of a U1 snRNP particle showing the organization of the Sm proteins, the U1 specific proteins U1-70K, U1-A, and U1-C, and the U1 snRNA (adapted from Stark et al., 2001).

It has been shown by Far Western analysis that SC35 bridges the U1 snRNP specific protein U1-70K and U2AF35 (Wu and Maniatis, 1993) thus serving to bridge the 5' and 3' ends of the intron (see Figure I-5).

The non-snRNP associated factor U2AF consisting of large (U2AF65) and small (U2AF35) subunits binds to the polypyrimidine tract and 3' splice site respectively. In addition, splicing factor SF1 binds to the branch point sequence, and members of the SR family of splicing factors constitute non-snRNP associated E complex components (Berglund et al., 1997; Das et al., 2000; Lee et al., 1993; Liu et al., 2001; Merendino et al., 1999; Michaud and Reed, 1991; Reed, 1989; Staknis and Reed, 1994; Wu et al., 1999; Zamore et al., 1992; Zorio and Blumenthal, 1999). The organization of these factors at the 3' splice site and the interactions between these factors has been extensively characterized (Figure I-5).

U2AF interacts with the branch binding protein SF1 to co-operatively promote SF1 recognition and binding to the branch region (Berglund et al., 1998). SF1 contains an N-terminal HK (hnRNP K) domain that is necessary and sufficient for BPS binding. The structure of the SF1-KH•BPS RNA complex demonstrates that SF1 contains a hydrophobic cleft that binds the RNA and results in the branch point adenosine being deeply buried within the protein (Liu et al., 2001). This recognition suggests that SF1 may facilitate subsequent formation of the pre-spliceosomal complex A by pre-bulging the branch adenosine for prior to U2 snRNA duplex formation. In yeast, the recognition of the branch region by the yeast splicing factor BBP (branch binding protein) involves U1 snRNP. These factors form a bridging interaction across the intron that involves U1 snRNP, the yeast U1 snRNP component Prp40, BBP, and Mud2p (Figure I-5A). The bridging interactions are conserved in the mammalian system in a structure involving U1 snRNP, SF1, and U2AF (Figure I-5B; Abovich and Rosbash, 1997). An interaction of the mammalian Prp40 like protein FBP11 (Formin Binding Protein) may promote an indirect U1 snRNP interaction with SF1 (Reed, 2000). There is some evidence that U2 snRNP may be weakly associated with the commitment complex; however, the association is not mediated by the pre-mRNA substrate and is independent of a functional branch point sequence (Champion-Arnaud et al., 1995; Das et al., 2000).

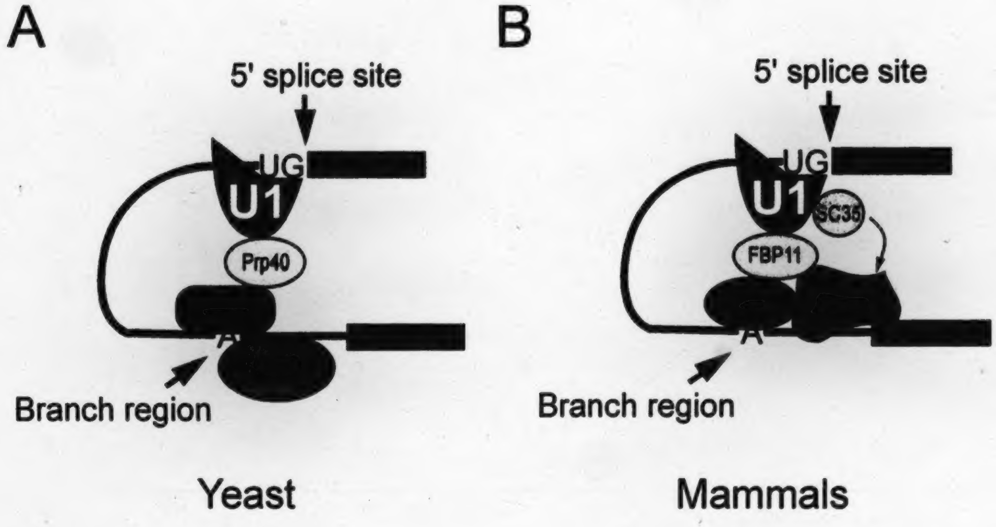


Figure I-5. Organization of commitment complex components. **(A)** The yeast commitment complex components U1 snRNP, branch binding protein (BBP), and Mud2p form a bridging interaction via the U1 snRNP protein Prp40 serving to bridge the branch region and 5' splice site. BBP interacts with the factor Mud2p. **(B)** The mammalian commitment complex components U1 snRNP, SF1, and U2AF form an analogous bridging interaction via FBP11 and SC35.

I-1.5. Formation of the pre-spliceosome commitment complex.

In yeast, two commitment complexes, CC1 and CC2 have been identified in U2 snRNP depleted extracts by native gel electrophoresis (Legrain et al. 1988; Seraphin and Rosbash, 1988; Zhang and Rosbash, 1999). The faster mobility complex CC1 is dependent on the presence of a 5' splice site and contains U1 snRNP. The lower mobility complex CC2 is dependent on the presence of a 5' splice site and a functional branch region and also contains U1 snRNP. In addition, several protein components of CC2 have been identified. Specifically, the splicing factor BBP (branch binding protein) has been shown to interact with the highly conserved yeast branch point sequence (Berglund et al., 1997) and U1 snRNP (Abovich and Rosbash, 1997; Zhang and Rosbash, 1999). Furthermore, Mud2p, the yeast homolog of U2AF65 (Zamore et al. 1992) has been found to interact with BBP which parallels the interactions of the mammalian splicing factor SF1 with U2AF65 (Berglund et al. 1998; Rūtz and Seraphin, 1999). These results suggest a similar structural organization of the yeast and mammalian commitment complex (Figure I-5). In mammals, only one commitment complex has been identified. Although the E complex in mammals has been well characterized, a pre-spliceosome complex that forms before E complex has not been identified. Previous work suggests that the E complex is the earliest mammalian pre-spliceosomal complex that commits the pre-mRNA to the spliceosome assembly pathway (Staknis and Reed, 1994).

I-1.6. The commitment complex component U2AF.

U2AF was first identified as an activity required for the stable recruitment of U2 snRNP to the branch region and extracts depleted of U2AF can be reconstituted for splicing by the addition of recombinant U2AF65 (Zamore and Green, 1989; Zamore et al., 1992). U2AF is a heterodimer consisting of 65 and 35 kDa subunits that bind the polypyrimidine tract and 3' splice site respectively (Figure I-6A).

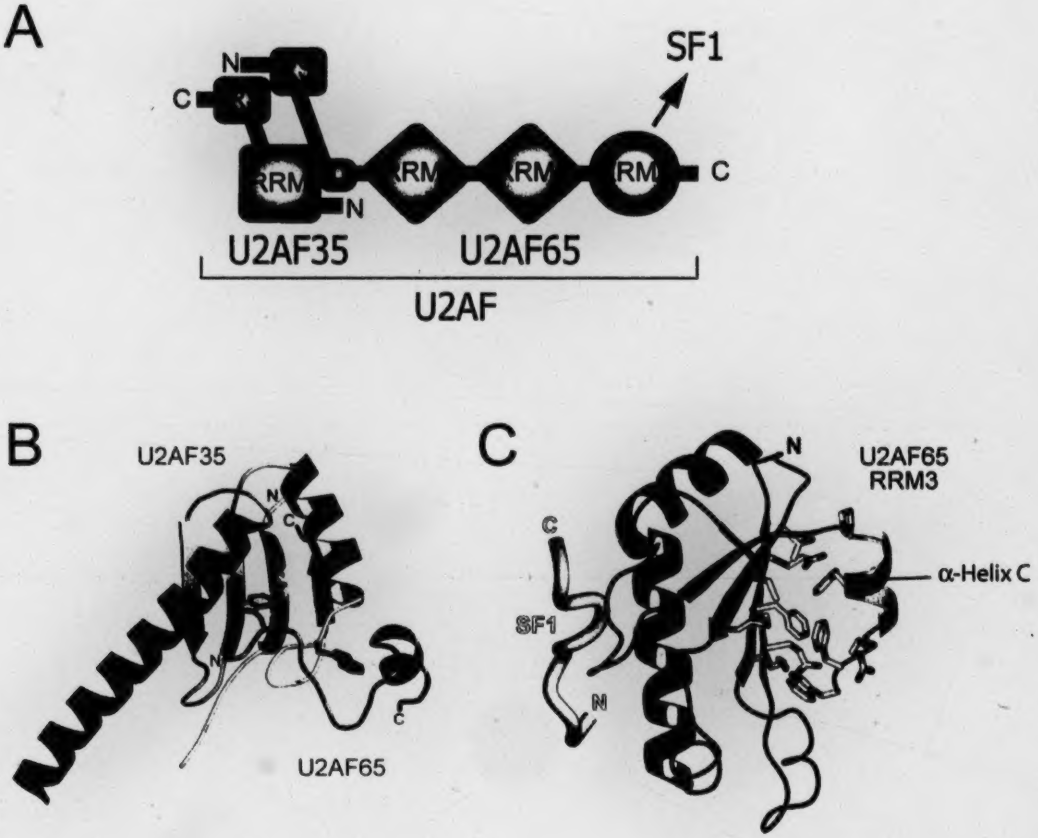


Figure I-6. The splicing factor U2AF. **(A)** U2AF is a heterodimer of large (65 kDa) and small (35 kDa) subunits. U2AF65 interacts with U2AF35 via the linker region (L). U2AF65 interacts with SF1 via RRM 3 (arrow). **(B)** Interface of U2AF35 with a U2AF65 peptide representing amino acids 85-112 (adapted from Kielkopf et al., 2001). **(C)** Interface of an N-terminal SF1 peptide amino acids 13-25 with U2AF65 RRM 3. Alpha-helix-C obstructs the RNA binding face of RRM3 (adapted from Selenko et al., 2003).

The large subunit, U2AF65, contains two canonical RNA recognition motif (RRM) domains required for RNA binding and one atypical RRM required for interaction with SF1 (Banerjee et al., 2003; Berglund et al., 1998; Selenko et al., 2003; Zamore et al., 1992). In addition, U2AF65 has an arginine/serine rich region (the RS domain) rich in basic residues and containing seven RS dipeptide repeats. The RS domain interacts with the branch point sequence and functions in U2 snRNP recruitment (Rudner et al., 1998; Valcarcel et al., 1996).

The small subunit, U2AF35, possesses an atypical RRM containing the site of dimerization with U2AF65 (Kielkopf et al., 2001) and a C-terminal RS domain (Figure I-6A) that can substitute functionally for the U2AF65 RS domain (Guth et al., 2001; Rudner et al., 1998). Specifically, the presence of one RS domain of U2AF, on either the large or small subunit, promotes high-affinity polypyrimidine tract binding activity consistent with redundant roles for the U2AF RS domains *in vivo* (Rudner et al., 1998). Recently, the structure of the interface of U2AF35 with a short peptide of U2AF65 consisting of amino acids 85-112 has been solved by X-ray crystallography that reveals two interlocking tryptophan residues (Figure I-6B).

Association of U2AF65 with the polypyrimidine tract is most likely mediated by two of the three RRM domains. Although it was initially reported that all three RRMs were required for tight RNA binding, recent experiments indicate that this may not be the case; RRMs 1 and 2 but not RRM 3 can be crosslinked to polypyrimidine RNA (Banerjee et al., 2003). The recent structure of the U2AF65•SF1 interface suggests that RRM 3 may not be an RNA binding module (Selenko et al., 2003; Zamore et al., 1992). Specifically, the NMR structure of the interface of an N-terminal SF1 peptide with U2AF RRM 3 demonstrates that the canonical RNA binding surface of RRM 3 is shielded by an α-helix which probably prevents RRM 3 binding to the polypyrimidine tract (Figure I-6C).

Recognition of the polypyrimidine tract by U2AF65 is central to splice site selection and spliceosomal formation for the major class of introns in higher eukaryotes; both the length and pyrimidine content of this sequence are critical determinants in the relative strength of competing 3' splice sites (Coolidge et al., 1997; Reed, 1989). Specific negative regulators of 3' splice site selection function as antagonists of

U2AF65•pyrimidine tract interaction and include the protein Sex-lethal (Sxl) which regulates female-specific alternative splicing of the *tra* pre-mRNA in *Drosophila* as well as the pyrimidine tract binding protein PTB which has been shown to modulate alternative splicing of the neural cell-specific c-*src* pre-mRNA (Chan and Black, 1997; Valcarcel et al., 1993).

I-2. Thesis overview.

I-2.1. Organization of the pre-mRNA substrate in the commitment complex.

A combination of genetic and biochemical experiments has elucidated a series of recognition events at both the 5' and 3' splice site of pre-mRNAs during spliceosome assembly. These include at least two separate recognition events at the 5' splice site within the commitment complex. Although the order of splice site recognition events is well established, the coordination of these events and especially the relationship between recognition of distal sequence elements is poorly understood. The structural organization of the pre-mRNA sequence elements within the mammalian pre-spliceosome E complex is the focus of this work.

We have used a combination of chemical and biochemical techniques to elucidate the proximity of the pre-mRNA 5' splice site, branch region, and 3' splice site within the commitment complex with respect to each other (Chapter 1). An established Fe-EDTA mediated hydroxyl radical probing technique, used to probe various interactions of the ribosome with both protein and tRNA factors (Joseph et al., 1997; Wilson and Noller, 1998), was employed to map the proximity of pre-mRNA sequence elements within pre-spliceosome complexes. N-terminal deletions of U2AF65 modified with hydroxyl radical probes were used to probe the structural organization of E complex factors at the 3' splice site (Chapter 2). These results have revealed the organization of the pre-mRNA substrate bound by 3' splice site factors and thus the disposition of U2AF65 on the pre-mRNA substrate.

It has been reported that there is a functional association between the 5' and 3' splice sites first established in E complex that involves U2AF and U1 snRNP (Michaud and Reed, 1993); therefore, we further investigated the early proximity of the pre-mRNA branch region within complexes formed in nuclear extracts depleted of U2AF (Chapter 3) or containing truncated U1 snRNA (Chapter 4). This work has allowed us to illuminate in more detail pre-spliceosome formation and to characterize the association of the 5' splice site, branch region, and 3' splice site during the assembly of pre-splicing factors on the pre-mRNA. Collectively, these results provide further insight into the formation of the pre-spliceosome commitment complex and organization of pre-mRNA substrates within these complexes.

I-3. References.

Abovich, N. and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89,** 403-412.

Anderson, K., and Moore, M.J. (2000). Bimolecular exon ligation by the human spliceosome bypasses early 3' splice site AG recognition and requires NTP hydrolysis. *RNA* **6**, 16-25.

Banerjee, H., Rahn, A., Davis, W., and Singh, R. (2003). Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* **9**, 88-99.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Nat. Acad. Sci. U.S.A.* **74**, 3171-3175.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.

Berglund, J.A, Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). In *The RNA World 2nd Ed.* (eds Gesteland, R.F., Cech, T.R., and Atkins, J.F.) 525-560 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Champion-Arnaud, P., Gozani, O., Palandjian, L., and Reed, R. (1995). Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol. Cell Biol.* **15**, 5750-5756.

Chan, R.C., and Black, D.L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.* **17**, 4667-4676.

Chaulk, S.G., and MacMillan, A.M. (2001). Separation of Spliceosome Assembly from Catalysis Using Caged pre-mRNA Substrates. *Angew. Chemie Int. Ed.* **40**, 2149-2152 / *Angew. Chemie* **113**, 2207-2210.

Chiara, M.D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *EMBO J.* **16**, 4746-4759.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, **1**, 1-8.

Chua, K., and Reed, R. (1999). The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* **402**, 207-10.

Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888-896.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell.* **5**, 779-787.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86-90.

Guth, S., Tange, T.O., Kellenberger, E., and Valcarcel, J. (2001). Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell Biol.* **21**, 7673-7681.

Jamison, S.F., and Garcia-Blanco, M.A. (1992). An ATP-independent U2 small nuclear ribonucleoprotein particle/precursor mRNA complex requires both splice sites and the polypyrimidine tract. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5482-5486.

Jamison, S.F., Crow, A., and Garcia-Blanco, M.A. (1992). The spliceosome assembly pathway in mammalian extracts. *Mol. Cell Biol.* **12**, 4279-4287.

Joseph, S., Weiser, B., and Noller, H.F. (1997). Mapping the inside of the ribosome with an RNA helical ruler. *Science* **278**, 1093-1098.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**, 595-605.

Konforti, B.B., Koziolkiewicz, M.J., and Konarska, M.M. (1993). Disruption of base pairing between the 5' splice site and the 5' end of U1 snRNA is required for spliceosome assembly. *Cell* **75**, 863-873.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.

Lee, C.G., Zamore, P.D., Green, M.R., and Hurwitz, J. (1993). RNA annealing activity is intrinsically associated with U2AF. *J. Biol. Chem.* **268**, 13472-1348.

Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell Biol.* **8**, 3755-3760.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.

MacMillan, A.M., Query, C.C., Allerson, C.A., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic Association of Proteins with the pre-mRNA Branch Region. *Genes Dev.* **8**, 3008-3020.

McSwiggen, J.A., and Cech, T.R. (1989). Stereochemistry of RNA cleavage by the Tetrahymena ribozyme and evidence that the chemical step is not rate-limiting. *Science* **244**, 679-683.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534-2546.

Michaud, S., and Reed, R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.* **7**, 1008-1020.

Moore, M.J. (2000). Intron recognition comes of AGe. *Nat. Struct. Biol.* **7**, 14-16.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364-368.

Moore, M.J., Query, C.C., and Sharp, P.A. (1993). In *The RNA World 1st Ed.* (eds Gesteland, R.F., and Atkins, J.F.) 303-347 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Newman, A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. *EMBO J.* **16**, 5797-5800.

Newby, M.I., and Greenbaum, N.L. (2002). Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* **9**, 958-965.

Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA* **9**, 968-980.

Nilsen, T.W. (1998). RNA-RNA Interactions in Nuclear Pre-mRNA Splicing. In *RNA Structure and Function*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), 279-307.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587-597.

Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P., and Krämer, A. (1998). Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *RNA* **4**, 551-565.

Rajagopal, J., Doudna, J.A., and Szostak, J.W. (1989). Stereochemical course of catalysis by the Tetrahymena ribozyme. *Science* **244**, 692-694.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113-2123.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**, 340-345.

Rossi, F., Forne, T., Antoine, E., Tazi, J., Brunel, C., and Cathala, G. (1996). Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 interaction. *J. Biol.Chem.* **271**, 23985-23991.

Rudner, D.Z., Breger, K.S., Kanaar, R., Adams, M.D., and Rio, D.C. (1998). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. *Mol. Cell Biol.* **18**, 4004-4011.

Rutz, B., and Seraphin, B. (1999). Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* **5**, 819-831.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF(65) and SF1/mBBP. *Mol. Cell* **11**, 965-976.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-358.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell Biol.* **14**, 7670-7682.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* ***92***, 315-326.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.

Tang, J., Abovich, N., Fleming, M., Serephin, B., and Rosbash, M. (1997). Identification and characterization of the yeast homologue of U1 snRNP-specific protein C. *EMBO J.* **16**, 4082-4091.

Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature **362**, 171-175.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706-1709.

Wilson, K.S., and Noller, H.F. (1998). Molecular movement inside the translational engine. *Cell* **92**, 131-139.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* ***402***, 832-835.

Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. USA* **86**, 9243-9247.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-838.

Chapter 1.

Early organization of pre-mRNA during spliceosome assembly.

Chapter 1. Early organization of pre-mRNA during spliceosome assembly (1).

1-1. Introduction.

1-1.1. Spliceosome assembly.

Intron excision from precursor mRNAs (pre-mRNAs) in eukaryotes requires the juxtaposition of reactive functionalities within the substrate at the heart of the spliceosome where the two chemical steps of splicing occur. Although a series of interactions between pre-mRNAs, pre-spliceosomal and spliceosomal factors is well established, the molecular mechanisms of assembly of the splicing machinery as well as the temporal basis for organization of the substrate for splicing remain poorly understood. Assembly of the spliceosome active site on the pre-mRNA proceeds through several stages and is directed by conserved sequences at the splice sites and within the intron as well as other sequences in the pre-mRNA. The current model for formation of the spliceosome active site involves an ordered, stepwise assembly of discrete snRNP particles on the pre-mRNA substrate.

In the canonical pathway (Burge et al., 1999; Kramer, 1996; Staley and Guthrie, 1998), commitment of a pre-mRNA to splicing involves the ATP independent formation of the early (E) or commitment complex on the RNA substrate. In mammals, this complex includes U1 snRNP tightly associated with the 5' splice site, loosely associated U2 snRNP (Das et al., 2000), as well as non-snRNP protein factors. These proteins include the heterodimer U2AF, containing large and small subunits, which binds to the polypyrimidine tract and 3' splice site, the branch binding protein SF1 which recognizes the branch region, and members of the SR protein family (Berglund et al., 1997; Merendino et al., 1999; Staknis and Reed, 1994; Wu et al., 1999; Zamore, et al., 1992; Zorio and Blumenthal, 1999). Formation of the mammalian E complex may be visualized by native agarose gel electrophoresis (Das and Reed, 1999); higher order A, B, and C splicing complexes may be separated from the non-specific H complex by native polyacrylamide gel electrophoresis (Konarska and Sharp, 1986). The A complex is

---

[1] Adapted from Kent, O.A., and MacMillan, A.M. (2002). *Nat. Struct. Biol.* **9**, 576-581.

formed in part by the stable, ATP dependent, association of U2 snRNP with the pre-mRNA: a duplex formed between U2 snRNA and the pre-mRNA branch region bulges out the branch adenosine specifying it as the nucleophile for the first transesterification (Query et al., 1994). Association of the U4/5/6 tri-snRNP with complex A produces complex B, which undergoes a series of rearrangements to yield complex C, the mature spliceosome. These rearrangements include displacement of U1 snRNP at the 5' splice site by U6 snRNP, the disruption of U4/U6 snRNA base-pairing, and the formation of a U2/U6 snRNA structure which is believed to form the active site of the spliceosome (Burge et al., 1999; Staley and Guthrie, 1998).

1-1.2. Intron structure within pre-spliceosomes.

In the mature spliceosome, the positioning of the splice sites relative to one another involves a number of RNA•RNA interactions. For example, there is a base-pairing between U6 snRNA and nucleotides proximal to the 5' splice site, U2 snRNA is base-paired to the pre-mRNA branch region, and U6 and U2 snRNA form a heteroduplex at the catalytic center of the spliceosome (Staley and Guthrie, 1998). It appears likely that the 5' and 3' exons are held in close proximity by interactions with U5 snRNA (Newman, 1997). Furthermore, there is a base-pairing interaction between the first and last nucleotides of the intron, required for the second step of splicing, which is conserved from yeast to mammals (Deirdre et al., 1997; Parker and Siliciano, 1993). Thus, in the spliceosome, the pre-mRNA substrate is structured to position the splice sites and branch region near to one another.

A variety of cross-intron protein•protein bridging interactions have been suggested in pre-spliceosomal complexes in both yeast and mammals. For example, both U2AF35 and the U1 snRNP associated protein U1 70K have been shown to interact with the SR protein SC35 by Far Western analysis (Wu and Maniatis, 1993). Similarly, it has been proposed that bridging interactions between U1 snRNP, the yeast branch binding protein (yBBP), and Mud2, the yeast homolog of U2AF, are conserved in the mammalian system in a structure involving U1 snRNP, mBBP (SF1), U2AF and possibly FBP11 the metazoan counterpart of yeast Prp40 (Abovich and Rosbash, 1997; Reed, 2000). It is unclear to what extent these interactions structure the pre-mRNA substrate for subsequent assembly of the spliceosome and organization of the substrate at the active site. Indeed, the current model of spliceosome assembly suggests essentially independent recognition of the 5' splice site and branch region by U1 and U2 snRNP respectively.

In order to determine the temporal basis for substrate alignment at the catalytic site of the spliceosome, we have used a directed hydroxyl radical probe (Fe-BABE), site-specifically attached to a pre-mRNA substrate, to map the structure of the pre-mRNA at early stages of spliceosome assembly. These studies indicate an early organization and proximity of conserved pre-mRNA sequences during spliceosome assembly/recruitment and suggest a mechanism for the formation of the final active site of the mature spliceosome. Our data show that all of the pre-mRNA sequences essential to the two steps of splicing are within close proximity (10-20 Å) at the earliest stages of spliceosomal assembly. We have demonstrated that the observed proximity depends on the activity of specific splicing factors including U1 snRNP and members of the SR protein class. Furthermore, this structure is dependent on the presence of a pyrimidine tract but not a functional branch sequence suggesting distinct roles for these sequences in the pre-mRNA substrate during the course of spliceosome assembly.

1-2. Results and Discussion.

1-2.1. Synthesis and characterization of Fe-BABE modified pre-mRNA.

Hydroxyl radicals generated from probes covalently attached to RNA or protein have been useful for examining both intra- and intermolecular contacts in RNA and RNA-protein complexes (Han and Dervan, 1994; Wilson and Noller, 1998). Diffusable radicals produced from a tethered Fe-EDTA moiety are excellent probes of local structure since they are only capable of cleaving the phosphodiester backbone within ~ 10-20 Å from their site of generation. We prepared pre-mRNAs containing a single tethered Fe-EDTA proximal to the branch region (position 148) by a combination of chemical (Figure 1-1A) and enzymatic synthesis (Figure 1-2A; Query et al., 1994; MacMillan et al. 1994; Moore and Sharp, 1992). A ten base oligomer containing the convertible nucleoside 4-chlorophenyl-uridine (Allerson and Verdine, 1995; MacMillan and Verdine, 1991) was chemically synthesized and deprotected with the bifunctional amine cystamine (see Appendix *II*) to yield an RNA containing a unique cytidine residue functionalized with a masked alkyl thiol tether. Upon treatment with DTT, this oligomer could be derivatized with iron bromoacetamidobenzyl-EDTA (Fe-BABE, Rana and Meares, 1990) in good yield to give the derivatized cytosine (Figures 1-1A and 1-1B). In order to prepare full-length pre-mRNAs modified with Fe-BABE, the thiol functionalized synthetic oligonucleotide was ligated to T7 RNA polymerase transcripts representing the 5' and 3' portions of the PIP85.B pre-mRNA substrate (Figure 1-2A; Query et al., 1994) in the presence of a DNA bridging oligonucleotide (MacMillan et al. 1994; Moore and Sharp, 1992). Reduction with DTT, followed by reaction with Fe-BABE, furnished a pre-mRNA with the desired modification (Figure 1-1A). The resulting RNA contained a single ^{32}P radiolabel one nucleotide 3' of the branch adenosine (Figure 1-2A) facilitating PAGE analysis of splicing and complex formation. We also prepared a pre-mRNA derivatized with Fe-BABE proximal to the 3' splice site (position 179) as well as two pre-mRNAs modified at position 148 in which elements required for splicing, the pyrimidine tract and the branch region, were mutated to disrupt spliceosome assembly.

We examined both splicing and ATP-dependent complex formation on the derivatized RNAs by incubating them in HeLa nuclear extract and then analyzing the results by denaturing and native PAGE (Figure 1-2B, data not shown). These experiments showed that the Fe-BABE modification at the branch site or 3' splice site had no effect on splicing or ATP dependent complex formation of the native pre-mRNA (Figure 1-2B; lanes 4-9, data not shown). As expected, either the deletion of the polypyrimidine tract or mutation to give a non-functional branch region severely impaired splicing of pre-mRNAs modified in this fashion (Figure 1-2B; lanes 10,11).

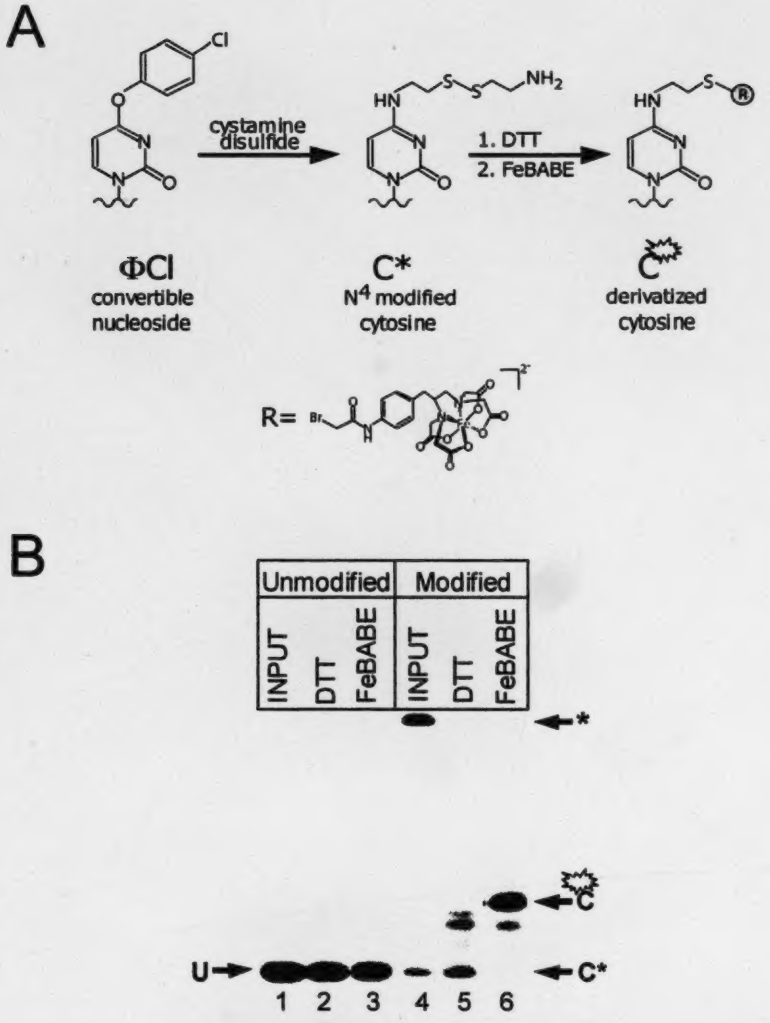


Figure 1-1. Synthesis of a branch oligonucleotide site-specifically modified with Fe-BABE. **(A)** Synthetic oligomers, representing the pre-mRNA branch region and containing a single chlorophenyl-uridine residue (ΦCl) were site-specifically functionalized by deprotection with the bifunctional amine cystamine (C*) and further modified by reduction and derivatization with Fe-BABE. **(B)** Unmodified (U) and modified (C) ^{32}P-labeled synthetic oligomers were reduced and derivatized with Fe-BABE and analyzed by denaturing PAGE. The asterisk indicates a disulfide cross-linked oligomer obtained following labeling.

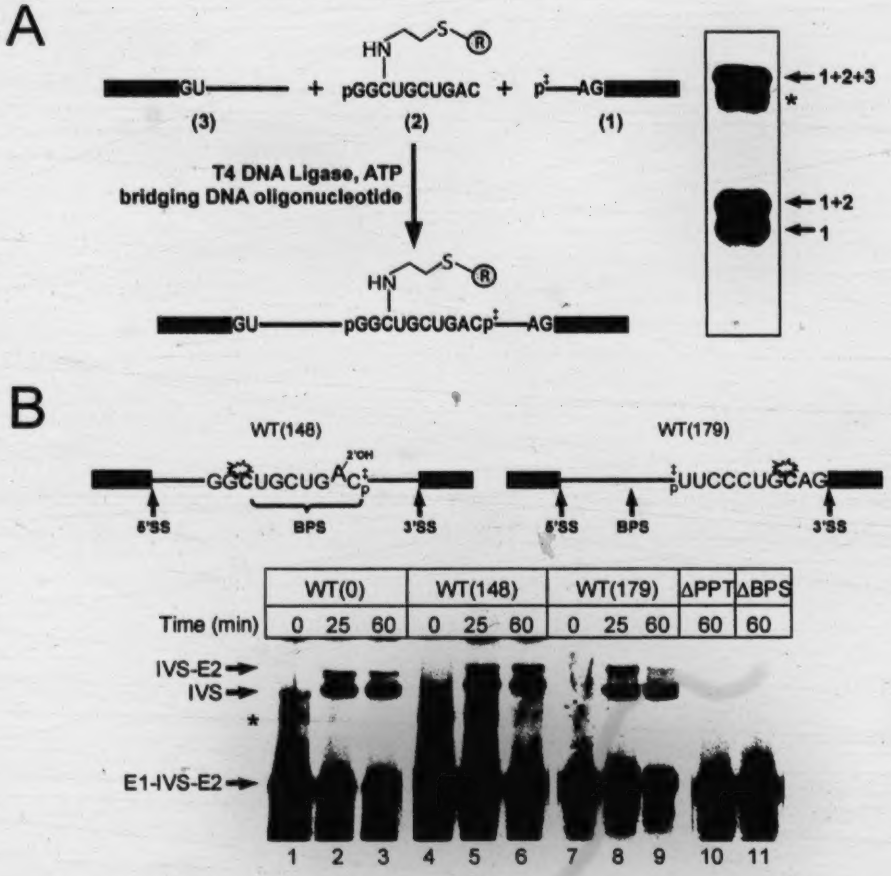


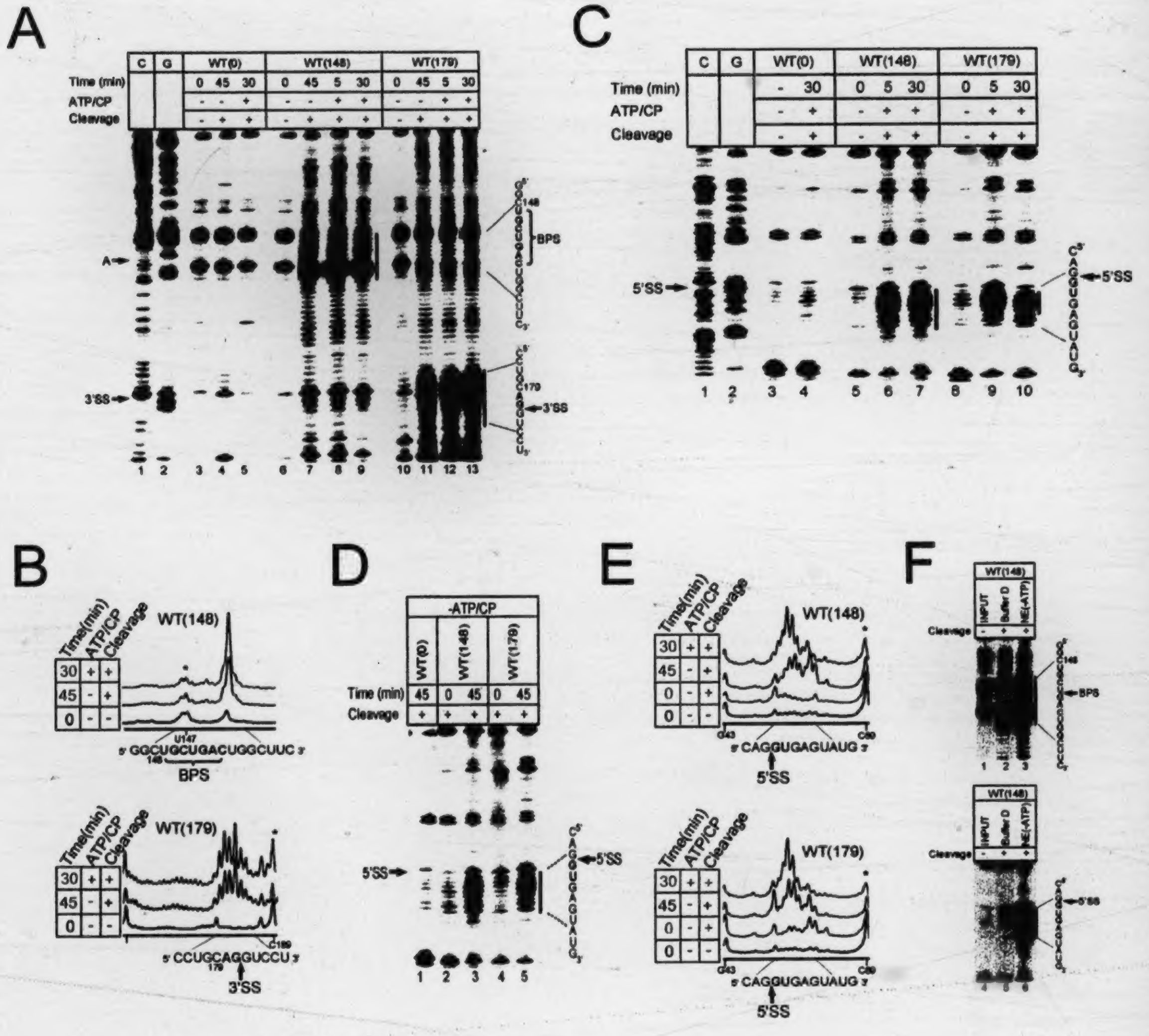
Figure 1-2. Synthesis and splicing of pre-mRNA site-specifically modified with Fe-BABE. **(A)** Functionalized 10-mers were ligated with transcription products representing the 3' and 5' portions of the PIP85.B pre-mRNA. Prior to ligation R= a sulfhydryl linked tether. Following ligation R=DTNB (see materials and methods) which is reduced and derivatized with Fe-BABE. Denaturing PAGE analysis of ligation reaction shown (right). Asterisk indicates degradation product. **(B)** (Top) Branch region and 3' splice site RNAs site specifically derivatized with Fe-BABE and containing a single ^{32}P radio-label (‡). (Bottom) Splicing of native and Fe-BABE derivatized pre-mRNAs. Lanes 1-3: wild type underivatized pre-mRNA (WT0); Lanes 4-11: Fe-BABE derivatized pre-mRNAs; lanes 4-6: Branch region modified pre-mRNA (WT148); lanes 7-9: 3' splice site modified pre-mRNA (WT179); lane10: modified pre-mRNA lacking a polypyrimidine tract (ΔPPT); lane 11: modified pre-mRNA with scrambled branch region (ΔBPS); IVS-E2: lariat intron intermediate, IVS: lariat product, E1-IVS-E2: pre-mRNA. Asterisk indicates a minor RNA product seen after incubation in nuclear extract that may result from poly-adenylation of the pre-mRNA transcript.

1-2.2. Mapping pre-mRNA structure in the pre-spliceosome E complex.

We investigated the structure of the pre-mRNA in HeLa nuclear extract under conditions associated with the formation of pre-spliceosomal (H,E,A/B) and spliceosomal (C) complexes by incubation for various times in the presence (A-C) or absence (H, E) of ATP followed by the initiation of Fe-BABE mediated cleavage. The results of Fe-BABE mediated cleavage were analyzed by RT/primer extension followed by denaturing PAGE (Stern et al., 1988). Primer extension using underivatized RNA as template (Figure 1-3A; lanes 3-5 and Figure 1-3C; lanes 3,4) generated a panel of reverse transcriptase control stops for comparison with cleavages observed in experiments involving Fe-BABE derivatized pre-mRNA (Figure 1-3).

As expected, significant cleavages, under a variety of conditions, were observed in the immediate vicinity of Fe-BABE modification for both the branch and 3' splice-site derivatized pre-mRNAs (as visualized by densitometric analysis and compared to control stops; Figure 1-3A; lanes 7-9 and lanes 11-13; Figure 1-3B). Interestingly, there was little change in the cleavage patterns observed over time periods associated with formation of the mature spliceosome; specifically, there appeared to be little communication between the branch region and 3' splice site as measured by cleavages observed using either derivatized RNA. This suggests that these regions are either positioned distal to one another throughout assembly of the spliceosome or that they are occluded from one another by bound protein or RNA components of the pre-spliceosome and spliceosome.

Strong Fe-BABE dependent cleavages were observed under both A/B and C complex forming conditions, proximal to the 5' splice site (position 55), using both branch region and 3' splice site derivatized RNAs (Figure 1-3C; lanes 6, 7 and 9, 10; Figure 1-3E). This is consistent with the facts that the branch region and 5' splice site are juxtaposed at the time of the first step of chemistry and that the 5' exon and 3' splice site are juxtaposed at the time of the second step of splicing.



(Figure 1-3. Directed hydroxyl radical probing in the spliceosome and pre-spliceosome. Legend on page 35.)

***Figure 1-3. Directed hydroxyl radical probing in the spliceosome and pre-spliceosome*.** **(A)** Directed hydroxyl radical probing of pre-mRNA proximal to sites of Fe-BABE modification in HeLa nuclear extract in the presence and absence of ATP. RNAs were incubated under the indicated conditions at 30°C followed by Fe-BABE mediated cleavage and then analyzed by RT primer extension. Lanes 1,2: C, G dideoxy sequencing ladders; lanes 3-5: control, wild-type underivatized RNA (WT0); lanes 6-9: branch region modified pre-mRNA (WT148); lanes 10-13: 3' splice site derivatized pre-mRNA (WT179). **(B)** Densitometric analysis of radical probing of derivatized pre-mRNA proximal to sites of Fe-BABE modification. (Top) Analysis of WT148 pre-mRNA in the presence of ATP (green), absence of ATP (blue) and input RNA (black). (Bottom) Analysis of WT179 pre-mRNA in the presence of ATP (green), absence of ATP (blue) and input RNA (black). **(C)** Directed hydroxyl radical probing at the 5' splice site in HeLa nuclear extract in the presence of ATP. RNAs were incubated under the indicated conditions at 30°C followed by Fe-BABE mediated cleavage and then analyzed by RT primer extension. Lanes 1,2: C, G dideoxy sequencing ladders; lanes 3-4: control, wild-type underivatized RNA (WT0); lanes 5-7: branch region modified pre-mRNA (WT148); lanes 8-10: 3' splice site derivatized pre-mRNA (WT179). **(D)** Directed hydroxyl radical probing at the 5' splice site in HeLa nuclear extract on ice (lanes 2 and 4) or at 30°C for 45 minutes (lanes 1, 3, and 5) in the absence of ATP. RNAs were incubated in ATP depleted HeLa nuclear extract on ice or at 30°C for 45 minutes followed by Fe-BABE mediated cleavage and then analyzed by RT primer extension. Lane 1: control, wild-type underivatized RNA (WT0); lanes 2 and 3: branch region modified pre-mRNA (WT148); lanes 4 and 5: 3' splice site derivatized pre-mRNA (WT179). **(E)** Densitometric analysis of radical probing of derivatized pre-mRNA at the 5' splice site. (Top) Analysis of WT148 pre-mRNA in the presence of ATP (green), absence of ATP (blue and red) and input RNA (black). (Bottom) Analysis of WT179 pre-mRNA in the presence of ATP (green), absence of ATP (blue and red) and input RNA (black). Individual graphs were normalized to a control band indicated with an asterisk for direct comparison. **(F)** Directed hydroxyl radical probing of pre-mRNA in the absence of nuclear extract. WT148 pre-mRNA was probed in buffer D (lanes 2 and 5; 30 minutes at 30°C) and the cleavages generated at the BPS (top) and 5' splice site (bottom) compared to both input RNA (lanes 1 and 4) and reactions probed under E complex conditions (lanes 3 and 6). Solid vertical bars represent sites of significant Fe-BABE mediated cleavage determined as regions greater than control bands. The sequences of PIP85.B pre-mRNA around the BPS, 3' and 5' splice sites indicated. The location of the site specific modifications C148 and C179 are shown.

Intriguingly, cleavages at the 5' splice site were observed under E complex forming conditions using both branch and 3' splice site derivatized pre-mRNAs (Figure 1-3D). These cleavages were strong, being roughly equal in intensity as compared with those observed under C complex forming conditions (Figures 1-3C, 1-3D, and 1-3E) when the conserved pre-mRNA sequences are directly juxtaposed. This result indicates a high order of substrate structure under conditions where only U1 snRNP and a number of non-snRNP splicing factors are stably associated with the pre-mRNA. As a control experiment, we initiated cleavage in the presence of buffer but no extract and failed to note any significant cleavages at the 5' splice site (Figure 1-3F). It should be noted that hydroxyl radicals are not capable of detecting a separation of components greater than 40-50 Å (Whirl-Carrillo et al., 2002); however, the heterogeneity of length observed in pre-mRNA introns as well as the lack of conserved sequence elements (Burge et al., 1999) argues against an organization of intron structure in the absence of protein factors.

1-2.3. Factor and sequence requirements in pre-spliceosome assembly.

The appearance of cleavages at the 5' splice site, upon incubation in HeLa extract in the absence of ATP, using the branch derivatized RNAs could be correlated with the appearance of the commitment complex as assayed by native agarose gels (Figure 1-4A; lanes 1, 2 and Figure 1-4B; lanes 4, 5; Das and Reed, 1999). We investigated the specificity of the observed cleavages in the pre-spliceosome by examining both the factor and pre-mRNA sequence requirements for cleavage in the vicinity of the 5' splice site. Again, there was an excellent correspondence between formation of the commitment complex and appearance of 5' splice site cleavages (Figure 1-4).

We specifically depleted HeLa extract of U1 and U2 snRNP using biotinylated 2'-O-methyl anti-U1 and anti-U2 oligonucleotides (Blencowe et al., 1989; Blencowe and Lamond, 1999). Depletion of U1 snRNP significantly diminished but did not abrogate the observed cleavages (Figure 1-4B, lane 6) but depletion of U2 snRNP had no effect (Figure 1-4B; lane 7).

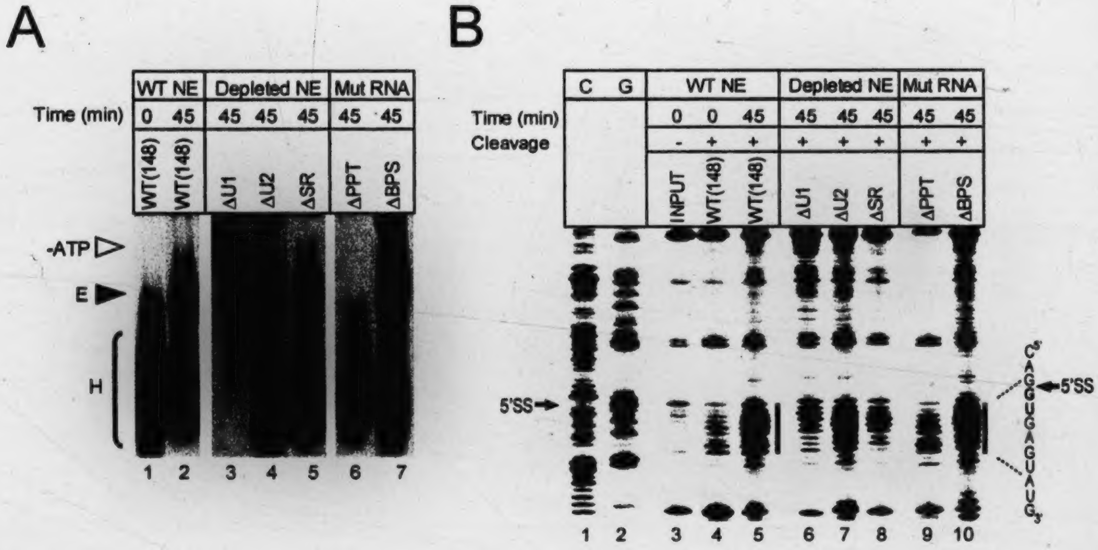


Figure 1-4. Factor and sequence dependence of E complex formation and Fe-BABE mediated cleavages. **(A)** Native agarose gel electrophoresis of E complex formation. E complex formation was analyzed on modified pre-mRNAs in wild type HeLa nuclear extract (WT NE), in depleted NE, and on pre-mRNA lacking a polypyrimidine tract (ΔPPT) and pre-mRNA with scrambled branch region (ΔBPS). Indicated are the H, E, and U1 independent (-ATP, Jamison and Garcia-Blanco, 1992) complexes. **(B)** Branch modified RNAs were incubated in ATP depleted HeLa nuclear extract on ice (lane 4) or at 30°C for 45 minutes (lanes 5-10) followed by Fe-BABE mediated cleavage and then analyzed by RT primer extension. Lanes 1, 2: C, G dideoxy sequencing ladders; Lanes 3-10: branch region modified RNA; lane 3: control, input RNA; lane 6: HeLa nuclear extract pre-treated with an anti-U1 snRNA 2'-O-methyl oligonucleotide; lane 7: HeLa nuclear extract pre-treated with an anti-U2 snRNA 2'-O-methyl oligonucleotide; lane 8: HeLa nuclear extract depleted of SR proteins; lane 9: pre-mRNA lacking a polypyrimidine tract; lane 10: pre-mRNA with a scrambled branch region. Solid vertical bars represent sites of significant Fe-BABE mediated cleavage. The sequence of PIP85.B pre-mRNA around the 5' splice site is indicated.

This is consistent with the cleavage at the 5' splice site reflecting formation of either the E complex or an E-like complex in which U1 snRNP is bound at the 5' splice site but does not require factors associated with higher order spliceosomal complexes.

Members of the SR protein family are constitutive non-snRNP associated splicing factors that associate with the pre-mRNA substrate at very early stages of spliceosome assembly and are required for transitions in the assembly pathway (Staknis and Reed, 1992; Roscigno and Garcia-Blanco, 1995). We depleted HeLa extract of SR proteins (Zahler et al., 1992) and observed that this treatment substantially diminished the 5' splice site cleavages (Figure 1-4B; lane 8). This observation may be explained generally as the requirement of SR proteins for formation of the commitment complex and specifically their role in bridging interactions between the branch region, 3' splice site and the 5' splice site (Wu and Maniatis, 1993). Indeed, the observation of some cleavages at the 5' splice site in the absence of U1 snRNP is consistent with the suggestion by Roth and co-workers that SR proteins alone are sufficient for bridging interactions across an intron (Stark et al., 1998).

We also examined complex formation on mutant RNAs lacking a polypyrimidine tract or with a scrambled branch region. Removal of the polypyrimidine tract from the pre-mRNA substrate severely decreased cleavages at the 5' splice site using pre-mRNAs derivatized in the branch region (Figure 1-4B; lane 9). This is consistent with the observation that the pre-mRNA lacking a pyrimidine tract does not assemble into E complex (Figure 1-4A; lane 6). In contrast, E complex formation and 5' splice site cleavage were still observed in substrates derivatized in an analogous position but lacking a functional branch sequence (Figure 1-4A; lane 7 and Figure 1-4B; lane 10). These results agree with the requirement of the polypyrimidine tract but not the branch sequence for assembly of the commitment complex (Champion-Arnaud et al., 1995). Although SF1 is believed to form part of the mammalian commitment complex, a weak dependence of E complex formation on the branch sequence is not unreasonable given the highly degenerate nature of this sequence in mammalian introns (Burge et al., 1999).

1-2.4. Implications for spliceosome assembly.

Our results suggest that pre-mRNAs are organized at a very early stage of spliceosome assembly such that the 5' splice site and branch region are directly proximal to one another (Figure 1-5A). The lack of an ATP requirement for all of the observed cleavages, combined with the depleted factor and pre-mRNA sequence requirements suggest that this organization corresponds with the formation of the splicing commitment complex. The closeness of the splice site•branch association observed here (~20 Å) is such that it likely directly templates the later stages of spliceosome assembly such as the formation of the U2 snRNA•branch region duplex as well as association of U6 snRNA with the 5' splice site (Figure 1-5B). In the present study, we also noted early proximation of the 5' and 3' splice sites; this is likely due to recognition of the 3' splice site CAG sequence by U2AF35 (itself bound as part of U2AF to the polypyrimidine tract proximal to the branch). While this structure is likely general to the pre-spliceosome, given the close distance between the branch region and 3' splice site in most introns, it is not obligate for spliceosome assembly since AG independent introns can undergo the first step of splicing in the absence of a 3' splice site (Anderson and Moore, 1997).

The current model of spliceosome assembly involves an ordered stepwise association of snRNP factors with the pre-mRNA substrate and includes essentially independent recognition of the 5' and 3' splice sites by U1 and U2 snRNP, respectively, followed by recruitment of the U5/U4/U6 tri-snRNP to the pre-mRNA. Recently, this view has been strongly challenged by Abelson and coworkers who argue that, in yeast, spliceosome assembly involves association of either a penta-snRNP (containing all of the spliceosomal snRNPs) or tetra-snRNP (lacking U1 snRNP) complex with the pre-mRNA substrate (Stevens et al., 2002). Isolated penta-snRNP particles were demonstrated to restore splicing activity to extracts depleted of endogenous RNAs; importantly, this reconstitution occurred with no exchange of U2 snRNP from the particle although a low amount of exchange of U1 snRNP was observed. These observations are consistent with a variety of reports that hint at the importance of higher order snRNP complexes in spliceosome assembly or a tighter coordination of the assembly process than previously recognized.

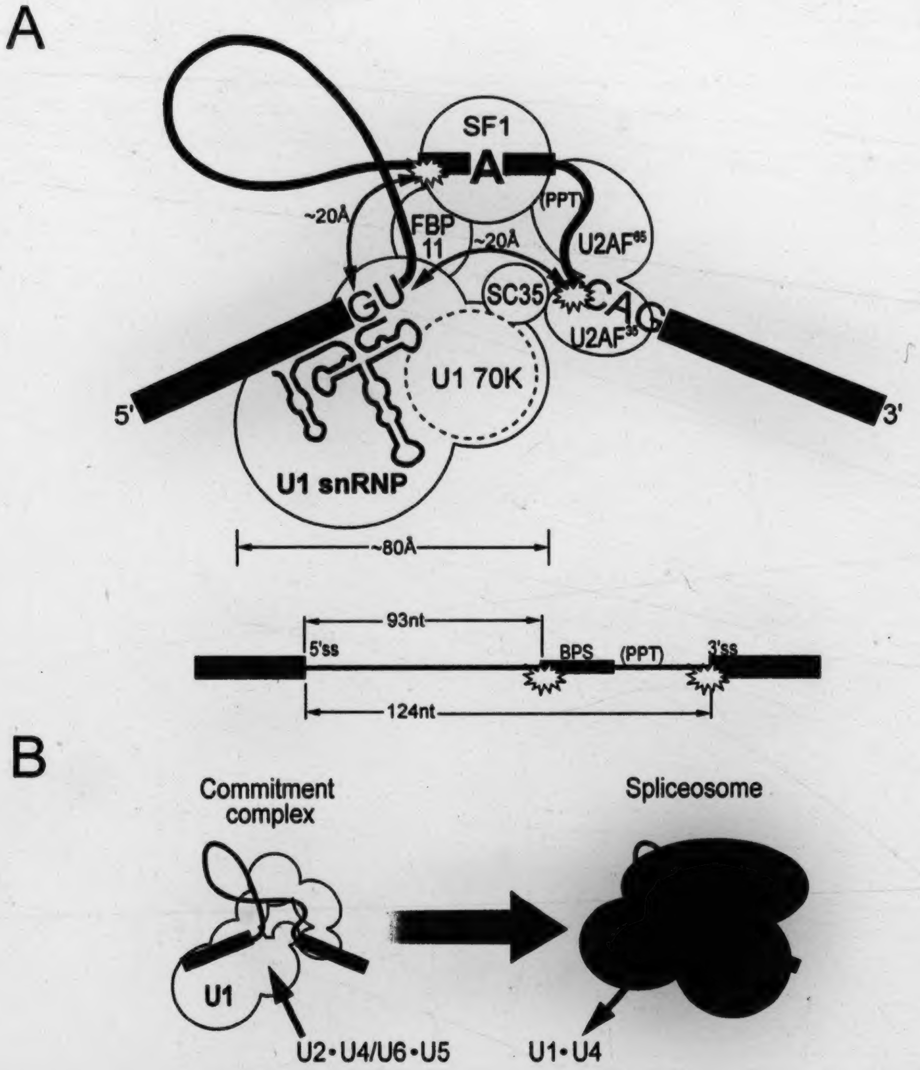


Figure 1-5. Early organization of pre-mRNA in spliceosome assembly. **(A)** (Top) Formation of the pre-spliceosomal commitment (E) complex involves various protein bridging interactions and results in the close proximation (~20 Å) of the branch region, 5' and 3' splice sites. This structure depends on the activity of U1 snRNP, members of the SR protein family, and the presence of a functional polypyrimidine tract (PPT) within the pre-mRNA substrate. The approximate size of U1 snRNP is noted for comparative purposes (Stark et al., 2001). (Bottom) The nucleotide distances of sequence elements in the PIP85.B pre-mRNA. **(B)** The structured pre-mRNA in the commitment complex templates the formation of the mature spliceosome. The mature spliceosome features association of U6 snRNA with the 5' splice site as well as U2 snRNP with the branch region. Formation of the U2-U6 heteroduplex at the catalytic center of the spliceosome is templated by the early proximation of the 5' splice site and branch region.

Given the striking conservation of the constitutive splicing machinery from yeast to mammals, it seems likely that mammalian spliceosome assembly is more of a coordinated process than previously thought.

The results reported in the present study are consistent with a model in which assembly of the spliceosome is essentially a concerted process involving coordinated recognition and organization of conserved sequences of the pre-mRNA substrate. Because U1 snRNP is tightly bound at the 5' splice site in complexes which can commit pre-mRNA substrates to splicing, we suggest that either U1 snRNP recruits a higher order snRNP complex to a structured pre-mRNA or that binding of U1 snRNP to the RNA is the means of recruitment of a U1 snRNP containing multi-snRNP complex to the pre-mRNA in a concerted recognition and organization of the sequences required for pre-mRNA splicing.

1-3. Materials and Methods.

1-3.1. Synthesis of Fe-BABE pre-mRNA substrates.

Synthesis of 4-chlorophenyl-uridine oligonucleotides. Oligonucleotides representing the PIP85.B pre-mRNA branch point sequence (5'-GG(ClϕU)UGCUGAC-3'), representing a scrambled branch point sequence (5'-GG(ClϕU)GUCGUAC-3'), and representing the PIP85.B 3' splice site (5'-UUCCCUG(ClϕU)AG-3') were synthesized containing 4-chlorophenyl-uridine (ClϕU). 4-chlorophenyl-uridine and modified RNA oligonucleotides were synthesized as outlined in Appendix *II*.

Plasmid preparation. The PIP85.B plasmid containing the PIP85.B pre-mRNA and an ampicillin resistance cassette was transformed into competent *E.coli* BL21 cells by incubation on ice followed by incubation in Luria Bertani (LB) at 37°C for 1 hour. Following transformation, cells were plated onto LB-agar plates containing 50 μg/mL ampicillin and grown overnight at 37°C. Isolated colonies were picked and grown in LB-amp broth overnight at 37°C. Overnights were centrifuged to pellet the cells, the supernatant was discarded and the cell pellet resuspended into lysis buffer (Qiagen).

Plasmid DNA was purified from cellular debris and bacterial DNA by mini-prep kit (Qiagen) following the manufacturer's protocol. Isolated purified plasmids were ethanol precipitated and resuspended in distilled water. Plasmid concentration was determined using a UV spectrophotometer at an absorbance of 260 nm.

PCR of transcription templates. Templates for transcription were made by PCR off the PIP85.B plasmid. The M13F primer (5'-CGC CAG GGT TTT CCC AGT CAC GAC-3') and CQ27 primer (5'-AGC TTG CAT GCA GAG ACC-3') were used to create a PCR product containing a T7 RNA polymerase promoter representing a transcription template for the full length PIP85.B pre-mRNA. The primer CQ36 (5'-TCG AGA CGA GCT GAC ATC-3') and M13F were used to create a PCR product for the transcription of a 5' piece RNA representing the 5' portion of the PIP85.B pre-mRNA. The primer CQ5 (5'-CCG GGA TCC TAA TAC GAC TCA CTA TAG GCT TCT TCT CTC TTT TTC CCT CAG GTC C-3') containing a T7 promoter sequence was used with CQ27 to create a transcription template for the transcription of a 3' piece RNA representing the 3' portion of the PIP85.B pre-mRNA. Analogously, the primer CQ32 (5'-CCT AAT ACG ACT CAC TAT AGG GAC GGA CAT GCA ATG CAA CTC AGG TCC TAC AC-3') was used with CQ27 to create a transcription template for the transcription of a 3' piece polypyrimidine tract knock-out RNA representing the 3' portion of the PIP85.B pre-mRNA. PCR reactions contained approximately 100-400 ng plasmid DNA and 100 μM of each primer. Plasmid and primers were incubated in reactions that contained 10 mM Tris-HCl pH 9, 1.5 mM $MgCl_2$, 50 mM KCl, 1 mM of each dNTP and 2 units TAQ DNA polymerase. Reactions were layered with mineral oil and subjected to a standard touch down PCR cycle on an MJ Research Mini-cycler. Following PCR, templates were extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. PCR products were resuspended into 50 μL distilled water and stored at -20°C. The efficiency of PCR reactions was determined by running an aliquot of PCR product on a 2% (w/v) horizontal agarose gel with 1×TBE running buffer.

Transcription. RNAs were transcribed from PCR products derived from the PIP85.B plasmid. Full length pre-mRNA (13/27), 5' piece (13/36) and 3' piece (5/27) were transcribed from the indicated PCR products. Standard transcription reactions contained 40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM NTPs, 10 mM DTT, 100-400 ng PCR derived template, and 10 units RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, centrifuged to remove insoluble inorganic phosphate accumulated during the reaction, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. In order to create a 5' piece RNA with homogenous 3' ends a low magnesium transcription procedure was employed. Low magnesium transcriptions contained 40 mM Tris-HCl pH 7.9, 3 mM $MgCl_2$, 2 mM spermidine, 25 mM NaCl, 1 mM NTPs, 4 mM DTT, 100-400 ng PCR derived template, and 8 units RNA polymerase. Transcriptions were incubated at 37°C for 1.5 hours, centrifuged to remove insoluble inorganic phosphate accumulated during the reaction, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated reactions were resuspended into loading dye/water (1:3 v/v) and loaded onto 8-15% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. RNAs were visualized by UV shadowing and excised from the gel. Gel slices were crushed, resuspended in 0.3 M sodium acetate containing 10% (v/v) phenol, and extracted at 37°C overnight. Following extraction, RNAs were filtered through spin columns to remove gel pieces, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended with distilled water, and quantified by UV absorbance at 260 nm.

5'-end labeling. Transcribed RNAs were 5' dephosphorylated using calf intestinal alkaline phosphatase (CIAP, New England Biolabs). CIAP reactions contained 100-500 pmol RNA, 1×NEB3 buffer, 0.5 units RNase Inhibitor (Roche), and 50 units CIAP. Reactions were incubated at 37°C for 30 minutes, made 0.3 M in sodium acetate, extracted with phenol/chloroform/isoamyl alcohol twice, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended in 66 mM Tris-HCl, pH 7.6, 6.6 mM $MgCl_2$, 10 mM DTT, 1 mM ATP or 2-10 μM γ-[^{32}P]-ATP, and 5 units T4 polynucleotide kinase (Invitrogen). Labeling reactions were incubated at 37°C for 30 minutes, made 0.3 M in sodium acetate, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. For RNAs used in 3 piece ligation reactions, all three RNAs and the bridging DNA oligonucleotide were added to the kinase reaction prior to phenol/chloroform extraction and ethanol precipitated together.

RNA three-piece ligation. Ligation reactions were performed on a 100 pmol scale of 5'-^{32}P labeled RNA. Reactions contained 100 pmol of 5'-^{32}P labeled 3' piece, 200 pmol of N4-modified cytosine containing 10-mer oligonucleotide 5' labeled with cold ATP, 200 pmol of 5' piece with homogenous 3' ends, and 110 pmol of synthetic bridging DNA oligonucleotide. The ethanol precipitated RNA from the kinase reaction described above was resuspended into 31.5 μL distilled water containing 0.5 mM DTT, 66 mM Tris-HCl, pH 7.6, and 6.6 mM $MgCl_2$. The dissolved RNA/DNA mix was heated to 98°C for 30 seconds and then incubated at 55°C for 15 minutes to fully denature the RNA and reduce the disulfide containing 10-mer. The reaction was cooled on ice, followed by addition of 1.2 mM ATP, 0.2 units RNase inhibitor, 2.7 % (v/v) PVP40 (polyvinyl propylene), and 40 units DNA ligase (Roche). Ligations were incubated at 30°C for 3 hours, followed by the addition of 150 μL of DTNB (dithionitrobenzene, 5 mM, 40% DMSO) and incubation at 30°C for another hour. The ligation was quenched with 25 μL loading dye (0.4 mg/mL Bromo Phenol Blue/Xylene cyanole, 6.4 M urea, 0.4 M Tris-base, 0.4 M boric acid, and 100 mM Na_2EDTA), and immediately loaded onto an 8% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gel. Ligations were visualized by autoradiography, the bands corresponding to full length pre-mRNA excised, crushed, and eluted from the gel as described above. Ligated pre-mRNAs contain an internal ^{32}P label within the intron two nucleotides 3' of the branch point adenosine.

Formation of Fe-BABE. Bromoacetamidobenzyl-EDTA (BABE, Dojindo) was dissolved in DMSO to a final concentration of 50 mM. In order to chelate iron, BABE was incubated at a final concentration of 10 mM with 100 mM sodium acetate pH 6, and 10 mM $FeSO_4$. The chelation was incubated at room temperature for 1 hour, and then chased with 10 mM EDTA for 10 minutes on ice. Aliquots were removed and stored at -20°C for derivatization reactions.

Derivatization of N4-functionalized cytosine. N4-functionalized cytosine RNA oligomers were derivatized with Fe-BABE under reducing conditions. Similarly, pre-mRNA ligation products containing the N4-functionalized cytosine at the branch region or 3' splice site were derivatized in an analogous fashion. Ethanol precipitated RNA was resuspended in 2 mM DTT, 50 mM sodium acetate, and 2% (v/v) glycerol. The reaction was incubated at 55°C for 30 minutes to reduce the disulfide on the N4-functionalized cytosine and then chased with 10 mM Fe-BABE to a final concentration of 5 mM. The reaction was incubated at 37°C for 2 hours in the dark. Following derivatization, the reaction was applied to a pre-washed BIORAD P30 spin column to remove unincorporated Fe-BABE and then washed with distilled water. Pooled RNA washes were extracted with phenol/chloroform/isoamyl alcohol, and ethanol precipitated with 2 μg glycogen. Derivatized 10-mer oligonucleotides were resuspended in loading dye/water and fractionated on a 20% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gel pre-run with 1 M sodium acetate 1×TBE running buffer in the lower chamber. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager. Derivatized pre-mRNAs were resuspended in distilled water and stored at -20°C.

1-3.2. Pre-mRNA splicing and complex formation.

Pre-mRNA splicing. For analysis of splicing, pre-mRNAs (50-100×10^3 c.p.m.) derivatized with Fe-BABE were incubated in reactions containing 40% (v/v) HeLa nuclear extract (Accurate Chemicals) for various times. Splicing reactions contained 40% (v/v) nuclear extract, 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and 4 units RNase inhibitor. Reactions were incubated at 30°C for 0 to 60 minutes and stopped by incubation on ice. Following splicing, reactions were digested with Proteinase K. Six μL of PK mix containing 1 μg tRNA, 7 μg Proteinase K, and 40 % (v/v) stop solution (6% (w/v) SDS, 250 mM Tris-HCl, pH 8, 25 mM EDTA) was added to a 20 μL splicing reaction. Reactions were digested at 55°C for 20 minutes, and then extracted with phenol/chloroform/isoamyl alcohol, and ethanol precipitated.

Precipitated reactions were resuspended in loading dye and subjected to 15% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager. For analysis of E complex formation, pre-mRNAs ($50\text{-}100\times10^3$ c.p.m.) were incubated in reactions containing 25% (v/v) HeLa nuclear extract, 60 mM KCl, and 4 units RNase inhibitor. Reactions were incubated at 30°C for 45 to 120 minutes and stopped by the addition of 2 μL 97% (v/v) glycerol loading dye and incubated on ice. Five μL aliquots were fractionated on 1.2% (w/v) agarose, 0.5×TBE gels. Gels were poured in a horizontal gel apparatus, 0.5 cm thickness, 12×12 cm square and run at 75 volts limiting for approximately 8 hours. Gels were transferred to blotting paper, dried, and exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Preparation of factor depleted nuclear extracts. In order to deplete SR proteins from HeLa nuclear extract $MgCl_2$ was added to a final concentration of 20 mM followed by centrifugation at 8700 rpm (mini centrifuge, IEC micromax) for 30 minutes. The supernatant was removed and dialyzed into glycerol free buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT). All dialysis >1 mL were performed in 3000 MWCO dialysis tubing (Fisher Brand). For dialysis <1mL 3,500 MWCO Slide-A-Lyzer Dialysis cassettes (Pierce) were used. Extracts to be dialyzed were submersed in 1 L of appropriate buffer for 1 hour (×2) and then overnight at 4°C. U1 snRNP and U2 snRNP were specifically depleted from HeLa nuclear extract using an anti-sense affinity chromatography procedure. The nuclear extract was dialyzed against buffer MD[0.1] (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT, 0.2 mM EDTA, 10% (v/v) glycerol) for 3.5 hours. The extract was centrifuged at 3000 rpm for 10 minutes; the precipitate was discarded and the supernatant dialyzed against buffer MD[0.6] (20 mM HEPES, pH 7.9, 600 mM KCl, 0.5 mM DTT, 0.2 mM EDTA, 10% (v/v) glycerol) for 1.5 hours. Biotinylated-2'-O-methyl anti-sense oligonucleotides (Dharmacon) corresponding to the 5' arm of U1 snRNA (5'-GCC AGG UAA GUA UUU-3') and a previously determined 3' assessable region of U2 snRNA (5'-GGC CGA GAA GCG AUU U-3') were used for snRNP depletion. Anti-sense oligonucleotides were added to a final concentration of 3 μM with 1 mM ATP and 5 mM creatine phosphate to aliquots of nuclear extract and incubated at 30°C for 30 minutes. The blocked nuclear extract was then applied to pre-blocked streptavidin agarose beads. Streptavidin agarose beads were pre-blocked by adding 1/3 the bead volume to 2/3 wash buffer (40 mM Tris-HCl, pH 7.6, 3.0% (w/v) NaN_3 (w/v) 100 mM NaCl), 50 μg BSA, 5 μg glycogen, and 5 μg tRNA. The beads were rotated at 4°C for 20 minutes, followed by 3 washes with wash buffer. The nuclear extract blocked with anti-sense oligonucleotide was added to pre-blocked beads and rotated at 4°C for 45 minutes. The mix was centrifuged to pellet the beads bound to U1 or U2 snRNPs, and the depleted nuclear extracts were removed and dialyzed against buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT, 0.2 mM EDTA, 20% (v/v) glycerol). Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

1-33. Probing pre-mRNA structure in splicing complexes.

Affinity cleavage. Affinity cleavage reactions were performed in reactions containing 25% (v/v) HeLa nuclear extract that had been dialyzed into glycerol free buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT). Reactions for complexes A/B (5 minutes at 30°C) and reactions for complex C (30 minutes at 30°C) contained 2 mM $MgCl_2$, 20 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and 4 units RNase inhibitor. Reactions for H complex (incubated on ice) and reactions for E complex (45 minutes at 30°C) contained 20 mM KCl, and 4 units RNase inhibitor, but lacked ATP/CP and $MgCl_2$. Nuclear extract was depleted of endogenous ATP by prior incubation at room temperature for 5 minutes. Cleavage reactions were initiated by the addition of 0.1% (v/v) H_2O_2 and 10 mM ascorbic acid (final concentrations) and left on ice for 20 minutes. Reactions were quenched with 30 mM thiourea/1% (v/v) glycerol, digested with proteinase K, extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated.

RT primer extension. Primer extension reactions were performed using Super Script RT (Gibco BRL). For analysis of the branch region, polypyrimidine tract and 3' splice site the primer (5'-AGC TTG CAT GCA GAG AC-3') complementary to the 3' end of the 3' exon was used. For analysis of the 5' splice site the primer (5'-TCG AGA CGA GCT

GAC ATC-3') complementary to the region immediately 5' of the branch site was used. Pre-mRNAs isolated from splicing cleavage reactions were ethanol precipitated and resuspended in 3.5 μL 10% (v/v) glycerol/distilled water, 1 μL 4.5×hybridization buffer (225 mM K-HEPES, pH 7.0, 450 mM KCl), and 1 μL (2 μM) DNA primer. Reactions were centrifuged for 30 seconds to mix ingredients, incubated at 98°C for 1 minute and slow cooled in the heating block to 37°C. Reactions were centrifuged at high speed for 30 seconds to retrieve condensate on the inside of the tube and incubated on ice. For sequencing reactions, 1 μL of the appropriate 1.5 μM ddNTP stock was added to the inside of each tube. For extension reactions, 2 μL of extension mix (400 mM Tris-HCl, pH 8.5, 30 mM $MgCl_2$, 30 mM DTT, 17 mM dATP, 17 mM dGTP, 17 mM dTTP, 1 mM dCTP, 3 μCi α-[^{32}P]-dCTP, and 0.1 units Super Script reverse transcriptase) was added to each tube and pipetted to mix thoroughly. Reactions were incubated at 42°C for 30 minutes, then 1 μL chase mix (1 mM each dNTP) was added to the inside lid of each tube, centrifuged to mix, and further incubated at 42°C for 15 minutes. Reactions containing ^{32}P labeled RNA were digested with 2 μL 3 M NaOH at 55°C for 20 minutes, extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated. Following extension, reactions were subjected to 8% (19:1 acrylamide/bisacrylamide) denaturing PAGE, pre-run with 1×TBE and run with 0.5×TBE in the top chamber. Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager. Densitometric analysis was performed using Image Quant software.

1-4. References.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-412.

Allerson, C.R., and Verdine, G.L. (1995). Synthesis and biochemical evaluation of RNA containing an intrahelical disulfide crosslink. *Chem. Biol.* **2**, 667-675.

Anderson, K., and Moore, M.J. (1997). Bimolecular exon ligation by the human spliceosome. *Science* **276**, 1712-1716.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.

Blencowe, B.J., and Lamond, A.I. (1999). Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* **118**, 275-287.

Blencowe, B.J., Sproat, B.S., Ryder, U., Barabino, S., and Lamond, A.I. (1989). Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell* **59**, 531-539.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). In *The RNA World 2nd Ed.* (eds Gesteland, R.F., Cech, T.R., and Atkins, J.F.) 525-560 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Champion-Arnaud, P., Gozani, O., Palandjian, L., and Reed, R. (1995). Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol. Cell Biol.* **15**, 5750-5756.

Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**, 1504-1508.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell* **5**, 779-787.

Deirdre, A., Scadden, J., and Smith, C.W. (1995). Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14**, 3236-3246.

Han, H., and Dervan, P.B. (1994). Visualization of RNA tertiary structure by RNA-EDTA.Fe(II) autocleavage: analysis of tRNA(Phe) with uridine-EDTA.Fe(II) at position 47. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4955-4959.

Jamison, S.F., and Garcia-Blanco, M.A. (1992). An ATP-independent U2 small nuclear ribonucleoprotein particle/precursor mRNA complex requires both splice sites and the polypyrimidine tract. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5482-5486.

Konarska, M.M., and Sharp, P.A. (1986). Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* **46**, 845-855.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.

MacMillan, A.M., and Verdine, G.L. (1991). Engineering Tethered DNA Molecules by the Convertible Nucleoside Approach. *Tetrahedron* **47**, 2603-2616.

MacMillan, A.M., Query, C.C., Allerson, C.A., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic Association of Proteins with the pre-mRNA Branch Region. *Genes Dev.* **8**, 3008-3020.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

Moore, M.J., and Sharp, P.A. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**, 992-997.

Newman, A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. *EMBO J.* **16**, 5797-5800.

Parker, R., and Siliciano, P.G. (1993). Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**, 660-662.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587-597.

Rana, T., and Meares, C. (1990). Specific cleavage of a protein by an attached iron chelate. *J. Am. Chem. Soc.* **112**, 2457-2458.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**, 340-345.

Roscigno, R.F., and Garcia-Blanco, M.A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* **1**, 692-706.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell Biol.* **14**, 7670-7682.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.

Stark, J.M., Bazett-Jones, D.P., Herfort, M., and Roth, M.B. (1998). SR proteins are sufficient for exon bridging across an intron. *Proc. Natl. Acad. Sci. USA* **95**, 2163-2168.

Stern, S., Moazed, D., and Noller, H. (1988). Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481-489.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* **9**, 31-44.

Whirl-Carrillo, M., Gabashvili, I.S., Bada M., Banatao, D.R., and Altman, R.B (2002). Mining biochemical information: lessons taught by the ribosome. *RNA* **8**, 279-289.

Wilson, K.S., and Noller, H.F. (1998). Molecular movement inside the translational engine. *Cell* **92**, 131-139.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

Wu, S., Romfo, C. M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837-847.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-838.

Chapter 2.
Structuring of the 3' splice site by U2AF65.

Chapter 2. Structuring of the 3' splice site by U2AF65 (1).

2-1. Introduction.

2-1.1. Organization at the 3' splice site.

Assembly of the spliceosome on the pre-mRNA proceeds through the formation of several complexes and is directed by conserved sequences at the 5' and 3' splice sites as well as other sequences in the pre-mRNA; recognition of the 3' splice site is closely coupled to recognition of the proximal branch region and polypyrimidine tract within the intron. Regulation of this assembly process results in differential splice site usage and the resulting patterns of alternative splicing are a major source of proteome diversity in higher eukaryotes (Black, 2000; Maniatis and Tasic, 2002).

The heterodimer U2AF, containing large (U2AF65) and small (U2AF35) subunits, which binds to the polypyrimidine tract and 3' splice site, is a component of the ATP independent early (E) or commitment complex; a pre-spliceosomal complex that commits a pre-mRNA to the splicing pathway (Burge et al., 1999; Krämer, 1996; Statey and Guthrie, 1998). U2AF is at the center of a network of cross intron protein•protein bridging interactions in the commitment complex that have been extensively characterized in the mammalian system (see Introduction and Chapter 1). U2AF coordinates the complex process of 3' splice site recognition by virtue of indirect interactions with the branch through the branch binding protein SF1 and direct recognition of the conserved AG dinucleotide of the 3' splice site (Berglund et al., 1997; Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999).

---

[1] Adapted from Kent, O.A., Reayi, A., Foong, L., Chilibeck, K.A., and MacMillan, A.M. (2003). *J. Biol. Chem.* **278**, 50572-50577.

2-1.2. The splicing factor U2AF.

U2AF is a polypyrimidine tract binding factor required for early recognition of pre-mRNA substrates and recruitment of U2 snRNP to the branch region. Extracts depleted of U2AF can be reconstituted for splicing by the addition of recombinant U2AF65 (Zamore et al., 1992). U2AF65 contains three C-terminal RNA recognition motif (RRM) domains as well as an N-terminal region, commonly referred to as the RS domain, rich in basic residues and containing seven RS dipeptide repeats.

The role of the RS region in U2AF65 function has proven difficult to define. In U2AF depleted extracts, reconstitution of splicing activity by addition of U2AF65 required the presence of the RS domain (Zamore et al., 1992). Fusion of this domain to a heterologous polypyrimidine tract binding module also reconstituted splicing in depleted extracts (Valcarcel et al., 1993). The presence of an RS domain in U2AF is essential but redundant; studies in *Drosophila* have demonstrated that either the RS domain of U2AF65 or a similar domain in U2AF35 but not both was required for viability (Rudner et al., 1998). Green and coworkers initially suggested that the RS domain of U2AF65 was positioned over the branch region and functioned in the recruitment of U2 snRNP to the pre-mRNA and it was subsequently shown that U2AF65 possessed a non-specific RNA annealing activity (Lee et al., 1993; Valcarcel et al., 1993). In support of a model in which the RS region of U2AF65 was positioned near the branch region (Figure 2-1A), a crosslink could be formed between the RS domain and nucleotides directly proximal to the branch (Valcarcel et al., 1996). Subsequent to these studies, it has been determined that the branch binding protein SF1 interacts with the C-terminus of U2AF65 and it has been demonstrated that U2AF35 which interacts with N-terminal amino acids 85-112 of U2AF65 directly recognizes the 3' splice site (Merendino et al., 1999; Rain et al., 1998; Wu et al., 1999; Zorio and Blumenthal, 1999). These observations have been used to argue for an arrangement in which the C-terminus of U2AF65 is positioned proximal to the branch while the N-terminus is situated in the vicinity of the 3' splice site (Figure 2-1B, Berglund et al., 1998; Kielkopf et al., 2001; Selenko et al., 2003).

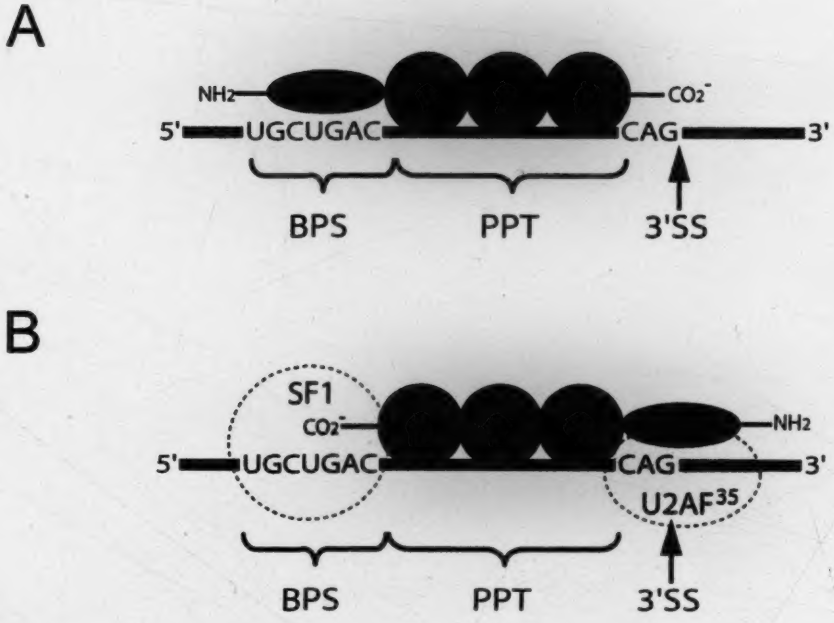


Figure 2-1. U2AF organization at the 3' splice site. **(A)** Model of U2AF65 binding the polypyrimidine tract to position the RS domain at the branch region. **(B)** Model of U2AF65 binding the polypyrimidine tract and interacting with SF1 at the branch region and U2AF35 at the 3' splice site, positions the RS domain at the 3' splice site.

Neither of these models accommodates all of the biochemical data and thus our current understanding of commitment complex structure at the 3' splice site is confused. Because of the central role of U2AF in organizing the assembly of factors at the 3' splice site and in promoting spliceosome assembly, we have undertaken a series of studies to determine the disposition of U2AF65 on the pre-mRNA.

We have investigated the positioning of the RS domain of U2AF65 with respect to bound RNA using a series of U2AF65 deletion mutants which have been site specifically modified at the N-terminus with the chemical nuclease Fe-EDTA. Our studies confirm not only that the RS domain is associated with the pre-mRNA branch region but also that the N-terminus of U2AF65 is close to the 3' splice site. This establishes the orientation of U2AF65 with respect to the RNA and indicates a pronounced bending of the RNA upon association of U2AF with the pre-mRNA. Structural data from X-ray studies of RRM•polypyrimidine complexes are consistent with both the proposed orientation as well as the RNA bending to juxtapose the branch and 3' splice site.

2-2. Results and Discussion.

2-2.1. Functionalization of U2AF65 with the chemical nuclease Fe-EDTA.

We decided to map the interaction of the N-terminus of U2AF65 with RNA by using a series of deletion mutants chemically modified at their N-terminus with Fe-EDTA as local probes of protein•RNA interaction. Preparation of this panel of probes was carried out by chemoselective reaction of an EDTA thioester (EDTA-3MPA) with the appropriate U2AF65 derived precursor modified to present an N-terminal cysteine residue following site-specific proteolysis (Figure 2-2A). Thioesters react specifically with N-terminal Cys since an initial thio-transesterification is followed by rapid rearrangement to yield a stable amide linkage (Figure 2-2B). Transient modification of internal Cys residues is reversed using an excess of DTT and the pH at which the modification is performed minimizes reaction with Lys side chains.

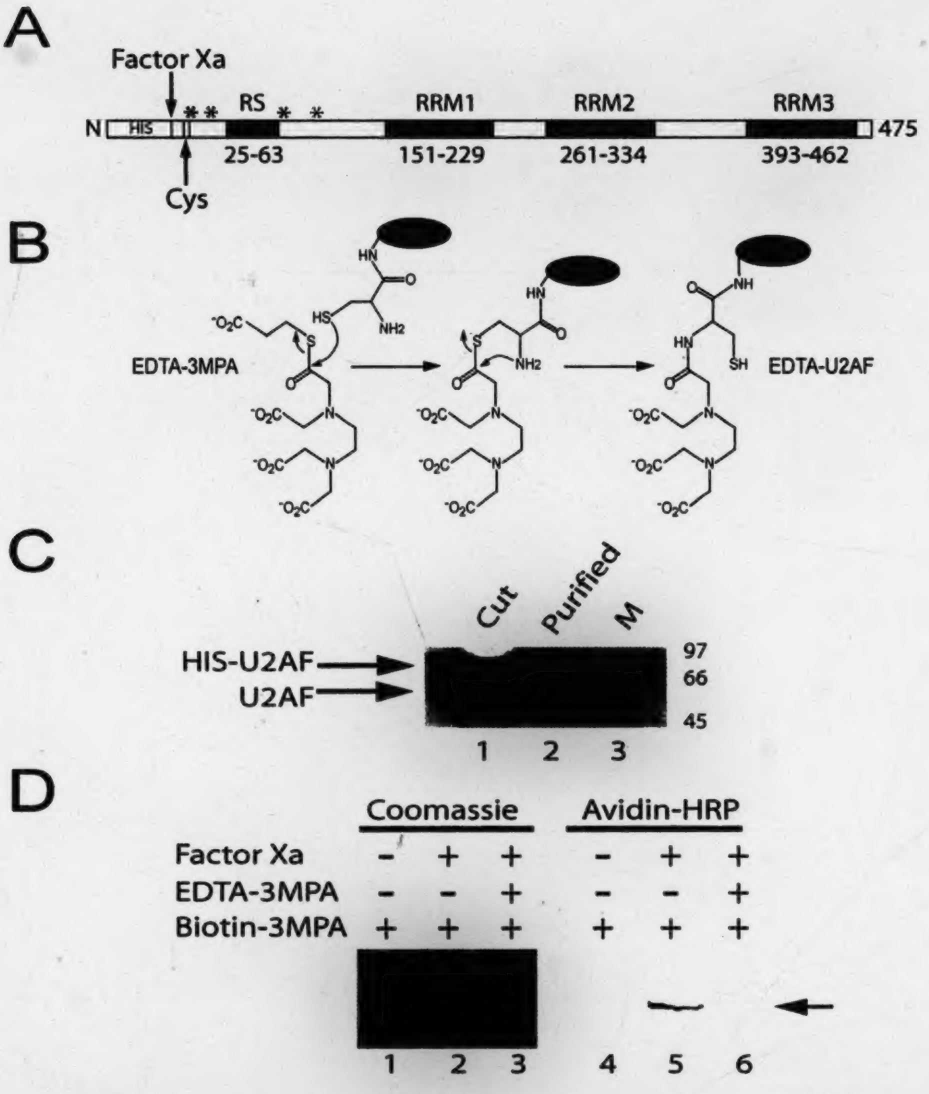


Figure 2-2. Modification of U2AF65 based probes with Fe-EDTA. **(A)** Domain structure of precursor proteins including N-terminus His6 tag followed by factor Xa recognition sequence. Cleavage with factor Xa exposes an N-terminal cysteine (Cys). Position of N-terminal deletions of U2AF65 indicated with colored asterisk (Blue WT, Green Δ1-14, Red Δ1-64, Grey Δ1-94). **(B)** Derivatization of N-terminal cysteine of U2AF65 using EDTA-3MPA yields a thioester which rearranges to give a stable amide linkage. **(C)** SDS PAGE of factor Xa cut U2AF65 (lane 1), and cut, derivatized and purified U2AF65-EDTA (lane 2). Molecular weight markers are shown (lane 3). **(D)** Site-specific derivatization of N-terminal cysteine with EDTA-3MPA. Uncut (lane 1 and 4), or cut and reacted with biotin-3MPA (lane 2 and 5), or cut and reacted with EDTA-3MPA (lane 3 and 6) and chased with biotin-3MPA and probed with avidin-HRP (lanes 4-6).

This strategy is based on work by Erlanson and Verdine who extended Kent's peptide ligation chemistry (Dawson et al., 1994) to the N-terminal functionalization of proteins with Fe-EDTA; in an elegant study, these workers were able to demonstrate NFAT mediated modulation of DNA binding by AP1 using patterns of DNA cleavage as a reporter of binding orientation (Erlanson et al., 1996). In the present work, a panel of four modified U2AF65 probes, representing successive N-terminal deletions, were prepared each containing an N-terminal His_6 tag and a factor Xa cleavage site followed by an introduced Cys residue (Figures 2-2A and 2-2B). Cleavage of partially purified protein yielded the desired precursor with an N-terminal Cys that was reacted with EDTA-3MPA and further purified to yield the functionalized probe (Figure 2-2C) that could then be activated by the addition of Fe^{2+}, ascorbic acid and hydrogen peroxide (Erlanson et al., 1996). In order to confirm that modification of the N-terminal Cys was specific, we performed modification/chase experiments with both EDTA-3MPA and the thioester biotin-3MPA since protein modification with biotin can be monitored by using an avidin-horseradish peroxidase conjugate (Figure 2-2D). These experiments showed that modification of U2AF65 was specific to the N-terminal Cys (in the presence of six internal Cys residues) and was essentially quantitative (Figure 2-2D; lanes 4-6).

2-2.2. Protection of the branch by the RS domain of U2AF65.

We chose a short (62 nucleotide) RNA consisting of the branch region polypyrimidine tract and 3' splice site to examine U2AF•RNA interactions. This RNA is a good model system for examination of 3' splice site recognition by U2AF65 and subsequent pre-spliceosome assembly; in HeLa nuclear extracts, a complex containing U2 snRNP stably associated with the branch region forms on this RNA in a U2AF65 dependant fashion (Query et al., 1997).

We examined the interaction of derivatized U2AF65 with RNA by native gel shift and RNase protection. Functionalized full length U2AF65 and the three deletion mutants bound the RNA similarly (Figure 2-3A; $K_D \sim 0.3\text{-}1\ \mu M$) with roughly equal affinity to that reported for U2AF65 alone (Rudner et al., 1998).

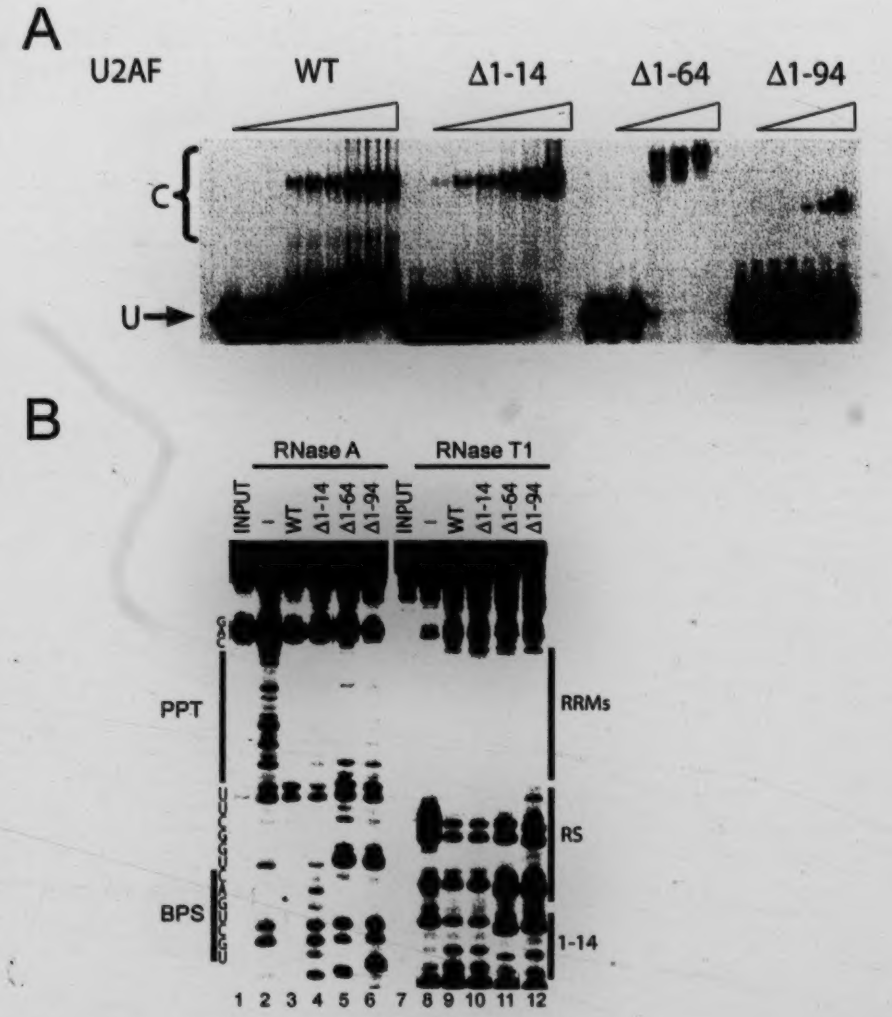


Figure 2-3. Mobility shift and RNase mapping of U2AF65•RNA interaction. **(A)** Gel shift of N-terminal deletion mutants of U2AF65 on wild type RNA. Unbound RNA (U) and bound U2AF65/RNA complexes (C) indicated. U2AF-N-terminal deletion mutants titrated from 0-10 μM final concentrations. All mutants bound with similar efficiency; $K_D \sim 0.3\text{-}1$ μM. **(B)** RNase A (lanes 2-6) and RNase T1 (lanes 8-12) footprinting of wild type RNA with N-terminal deletion mutants of U2AF65. Footprints from the RRMs, RS domain, and 1-14 amino acids indicated by vertical bars. Sequence around the branch point region (BPS) and 3' splice site (CAG) shown, polypyrimidine tract (PPT) indicated by a solid line. (-) No protein added.

Protection experiments with RNase A and RNase T1 showed an extended footprint on the RNA for full length U2AF65 from the 3' splice site through the polypyrimidine tract and including the branch point sequence (Figure 2-3B; lane 3). Deletion of the RS region resulted in a loss of protection in the vicinity of the branch region but in all three deletion constructs the polypyrimidine tract was protected consistent with the presence of the C-terminal RRMs (Figure 2-3B; lanes 3-6 and 9-12). Thus, the RS region either directly covers the branch region or indirectly modifies the RNA structure to protect it.

2-2.3. Mapping U2AF65•RNA interaction using tethered Fe-EDTA.

We investigated the structure of the 3' splice site RNA bound to the U2AF65 probes using 5' end-labeled RNA, initiating Fe-EDTA mediated cleavage under conditions where all of the RNA was bound, and analyzing the results by denaturing PAGE (Figure 2-4). Using full length U2AF65, we observed strong cleavage of the RNA in the vicinity of the branch region (Figure 2-4A; lane 4). A similar pattern of cleavage was observed using the U2AF65 construct derivatized directly adjacent to the RS domain (Δ1-14; Figure 2-4A; lane 5). These observations are consistent with the RS domain of U2AF65 directly contacting the branch region. Interestingly, using the probe derivatized adjacent to the RS domain (Δ1-14), some cleavage at the 3' splice site was also observed indicating that in the RNA•U2AF65 complex the branch region and 3' splice site are close to one another (Figure 2-4A; lane 5). When probes lacking the RS domain (Δ1-64, Δ1-94) were assayed, cleavage at the branch region was no longer seen; instead, strong patterns of cleavage proximal to the 3' splice site, roughly twenty-five nucleotides downstream were observed (Figure 2-4A; lanes 6 and 7). These cleavages were centered at the splice site (Δ1-64) and immediately 5' to the splice site (Δ1-94). These patterns of cleavage suggest an association of the N-terminal region of U2AF65 with the 3' splice site, consistent with the observation that the 3' splice site recognition factor U2AF35 interacts with amino acids 85-112 of U2AF65.

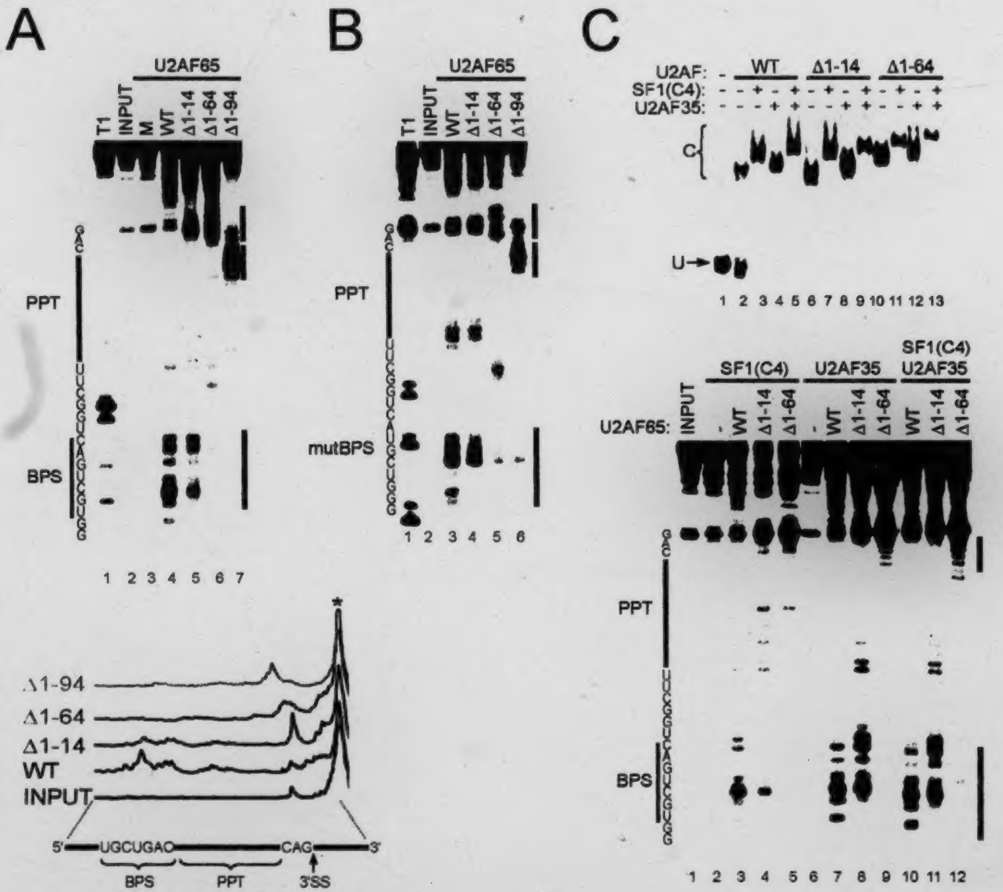


Figure 2-4. Directed hydroxyl radical cleavage of RNA by U2AF65-Fe-EDTA. **(A)** (top) Cleavage of wild type RNA with N-terminal deletion mutants of U2AF65. Reactions were incubated with U2AF65-Fe-EDTA (lanes 4-7) under cleavage conditions or with U2AF65-EDTA (lane 3, M) lacking iron. Input (lane 2) and RNase T1 (lane 1) shown. (bottom) Densitometric analysis of cleavage reactions (Blue WT, Green Δ1-14, Red Δ1-64, Grey Δ1-94). Individual graphs were normalized to full length RNA (asterisk). **(B)** Cleavage of mutant BPS RNA with N-terminal deletion mutants of U2AF65. Reactions were incubated with U2AF65-Fe-EDTA (lanes 3-6) under cleavages conditions. Input (lane 2) and RNase T1 (lane 1) shown. **(C)** Probing of structure in multi-protein complex. (top) Gel shifts of U2AF65 bound to RNA in the absence or presence of SF1-C4 and/or U2AF35 or both. Unbound RNA (U) and bound U2AF/RNA/factor complexes (C) indicated. (bottom) Cleavage of wild type RNA with N-terminal deletion mutants of U2AF65-Fe-EDTA in the absence or presence of SF1-C4 (lanes 3-5), U2AF35 (lanes 7-9), or both (lanes 10-12). Input (lane 1) and factor without U2AF65-Fe-EDTA (lanes 2 and 6) shown. (-) indicates absence of factor. Sites of significant cleavage indicated by vertical bars. Sequences around the branch point sequence (BPS) or the mutant branch point sequence (mutBPS) and 3' splice site (CAG) shown, polypyrimidine tract (PPT) indicated by a solid line.

The patterns of cleavage observed in these experiments are consistent with previous disparate observations (Merendino et al., 1999; Rain et al., 1998; Valcarcel et al., 1996; Wu et al., 1999; Zorio and Blumenthal, 1999) and suggest that they may be reconciled by a model in which U2AF65 bends the RNA to juxtapose the branch and 3' splice site. As well, the shift in cleavage pattern observed between the two constructs suggests a polarity of RNA recognition whereby the RNA is bound in a 3' to 5' sense N– to C– terminus on U2AF65 (Figure 2-5B).

We investigated the sequence dependence of RNA recognition by performing a series of cleavage experiments with the same panel of derivatized U2AF65 probes using an RNA containing a non-functional scrambled branch (Figure 2-4B). The probes bound this RNA with similar affinity (data not shown) and the cleavage patterns observed were similar to those seen using the consensus branch sequence (Figure 2-4A and 2-4B). Thus, structuring of the 3' splice site by U2AF65 is independent of the branch sequence; however, nucleotides 5' to the polypyrimidine tract do increase the affinity of U2AF65 for the RNA (data not shown) consistent with a non-specific interaction between the positively charged RS domain and the negatively charged RNA.

2-2.4. Structure of higher order complexes at the 3' splice site.

U2AF65 coordinates recognition of the 3' splice site by binding tightly to the polypyrimidine tract and through interactions with the branch binding protein SF1 and the 3' splice site recognition factor U2AF35. We repeated our analysis of U2AF65•RNA interaction using three of the U2AF65-based probes (Δ1-94 does not include part of the U2AF65/35 interface) in the presence of combinations of U2AF35 and C4, a deletion mutant of SF1 that is functionally equivalent to the full length protein (Rain et al., 1998). Using these factors we were able to form specific ternary and quaternary complexes on RNA mimicking the known assembly of factors at the 3' splice site as assayed by native gel shift (Figure 2-4C). The RNA cleavage patterns observed upon incubation of the U2AF65-based probes with SF1-C4 and U2AF35 alone or with both proteins together are very similar to those observed with U2AF65 alone (Figure 2-4C). We conclude that U2AF65 is the primary determinant of RNA structure at the 3' splice site within the commitment complex (Figure 2-5B).

2-2.5. Models of RNA recognition by polypyrimidine binding factors.

The splicing factor U2AF is the central organizing force in recognition of the 3' splice site in higher eukaryotes and as such plays a critical role in the regulation of gene expression. Our studies indicate that U2AF65 alone, bound to the 3' splice site region, bends the polypyrimidine tract to bring the 3' splice site and branch region close together. In order to refine a model of U2AF-RNA interaction we considered known RRM-polypyrimidine binding modes as determined by X-ray crystallography: specifically the structures of the Sxl and HuD proteins complexed to RNA (Figure 2-5A; Handa et al., 1999; Wang and Tanaka Hall, 2001). The *Drosophila* protein Sxl regulates female-specific alternative splicing of the *tra* pre-mRNA by binding to a specific polypyrimidine tract thus preventing the binding of U2AF and utilization of the proximal 3' splice site (Valcarcel et al., 1993). Human Hu proteins protect short-lived RNAs from degradation by binding to adenosine-uridine rich elements (AREs) in their 3' untranslated region.

The domain structure of Sxl includes tandem RRM domains that mediate binding to a seventeen nucleotide uridine rich sequence. The individual RRMs have a characteristic structure consisting of a four-stranded anti-parallel β-sheet buttressed by two α-helices. The 2.6 Å X-ray structure of Sxl bound to RNA reveals a V-shaped cleft formed by the β-sheets which are the principal determinants of RNA recognition; this cleft accommodates, in part, the RNA which is strongly kinked in the bound structure (Figure 2-5A). X-ray structures of the tandem ARE-binding RRMs of HuD bound to RNAs derived from the c-fos and TNF AREs show a remarkable similarity to each other and to the structure of Sxl bound to RNA although the bound RNAs differ in sequence at several positions; HuD also forms a cleft for the recognition of an extended RNA which is sharply kinked (Figure 2-5A). Given the similarity in consensus sequences recognized by U2AF, Sxl, and HuD, it seems likely that the general modes of RNA recognition are quite similar.

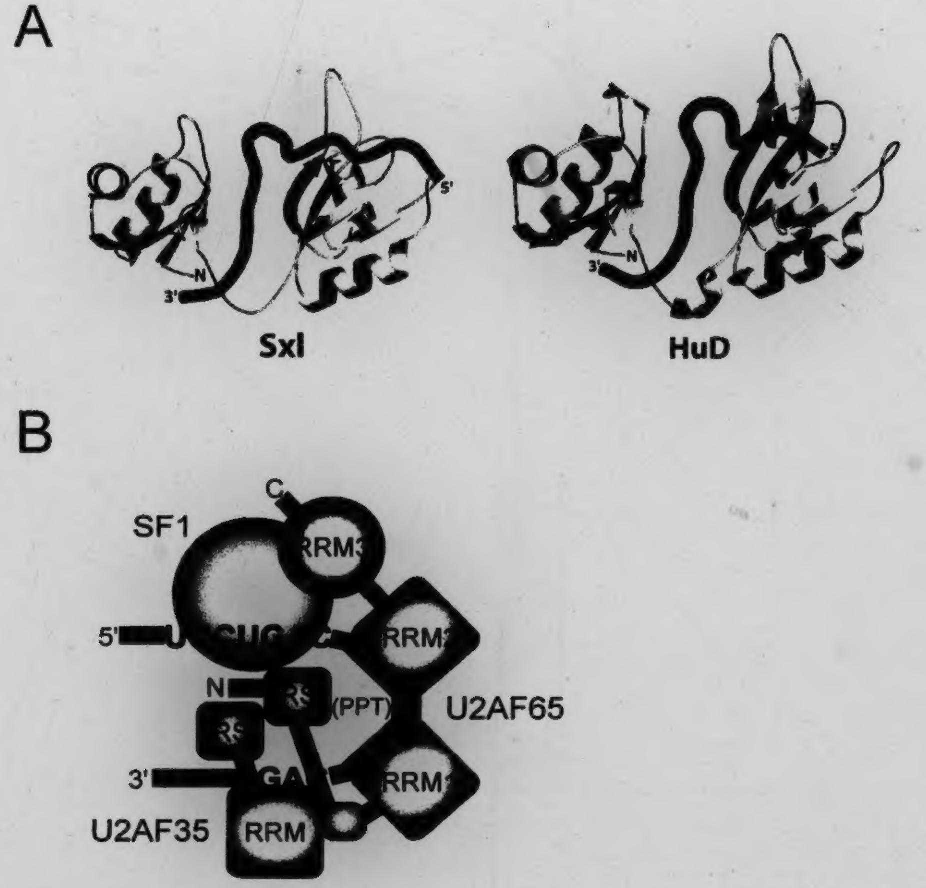


Figure 2-5. U2AF65 bends the polypyrimidine tract. **(A)** High-resolution structures of Sxl•RNA (left) and HuD•RNA (right) complexes (Handa et al., 1999; Wang and Tanaka Hall, 2001). Both polypyrimidine tract bending and C- to N-terminal orientation of the protein with 5' to 3' orientation of the RNA match the observed interaction of U2AF65 with the 3' splice site. **(B)** Orientation of U2AF65 at the 3' splice site structures the polypyrimidine tract to juxtapose the branch point region and 3' splice site. Interactions of U2AF65 with SF1 at the branch point region and U2AF35 at the 3' splice site in a multi-protein complex mimic the pre-spliceosome commitment complex.

Thus, association of U2AF65 with the pre-mRNA 3' splice site strongly bends the RNA as observed using the Fe-EDTA modified probes (Figure 2-5B).

U2AF65 recognizes a diverse array of polypyrimidine rich sequences with a typical length of twenty to forty nucleotides (Coolidge et al., 1997; Reed, 1989). In addition to direct recognition of the polypyrimidine tract, U2AF•RNA interaction is modulated by U2AF35 binding to the 3' splice site and indirectly by the strength of the branch region•SF1 interaction.

A seventeen nucleotide uridine rich consensus binding site for U2AF65 has been identified by SELEX (Singh et al., 1995); from both biochemical experiments and the high-resolution structural data for Sxl RNA interaction, it is expected that this sequence could be recognized by RRM1 and RRM2 of U2AF65. The recent NMR structure of the U2AF65•SF1 interface reveals that the canonical RNA binding surface of RRM3 of U2AF65 is blocked by an α-helix (Selenko et al., 2003). A similarly positioned helix is displaced in the association of the U1-A RRM with its cognate RNA. However, the atypical sequence of U2AF65 RRM3, lacking conserved RNA-interacting aromatics, the overall negative surface potential of the RRM, and the fact that RRM3 is not crosslinked to polypyrimidine RNA suggests that the functional role of this domain is in protein-protein recognition (Banerjee et al., 2003; Selenko et al., 2003). In contrast to the highly specific polypyrimidine tract recognition of Sxl, U2AF65 binds to tracts of varying length and sequence composition; long pyrimidine sequences may be recognized by a looping out of the RNA while differences in pyrimidine tract sequence may be accommodated by different registers of binding (Banerjee et al., 2003).

Recently, using pre-mRNAs site-specifically functionalized with Fe-EDTA, we have shown that both the branch and 3' splice site are in close proximity to the 5' splice site in the mammalian commitment complex. Consequently, we suggested that this arrangement of critical sequences templates further spliceosomal assembly and snRNA rearrangement (Kent and MacMillan, 2002). Structures of the first two RRMs of U2AF65, as well as the U2AF65•SF1 and U2AF65•U2AF35 interfaces and the SF1•branch region complex have now been determined (Ito et al., 1999; Kielkopf et al., 2001, Liu et al., 2001; Selenko et al., 2003). However, the spatial and structural organization of these factors at the 3' splice site is unclear. The studies described here suggest that the 3' splice site complex will have a relatively compact structure in which the key conserved intron sequences are positioned close to one another by virtue of their interaction with the U2AF65 scaffold. A more complete understanding of commitment complex assembly and function will involve high resolution structural analysis of higher order complexes containing these and other factors.

2-3. Materials and Methods.

2-3.1. Preparation of U2AF binding substrates.

RNA synthesis. RNAs for U2AF binding and footprinting reactions were transcribed from annealed synthetic DNA templates. The RNAs were based on the PIP85.B pre-mRNA sequence; RNAs contained a branch point sequence, polypyrimidine tract, and 3' splice site. The wild type RNA was transcribed from the DNA template (5'-TGT TGT GTA GGA CCT GAG GGA AAA AGA GAG AAG AAG CCA GTC AGC ACC CTC GAG ACG AGC CCT ATA GTG AGT CGT ATT A-3') with a short DNA oligonucleotide complimentary to the T7 promoter region (5'-TAT CAC TCA GCA TAA T-3'). Mutant RNAs with a scrambled branch point sequence (5'-TGT TGT GTA GGA CCT GAG GGA AAA AGA GAG AAG AAG CCA GTA CGA CCC CTC GAG ACG AGC CCT ATA GTG AGT CGT ATT A-3') and polypyrimidine knock-out sequence (5'-TGT TGT GTA GGA CCT GAG TTG CAT TGC ATG TCC GTC CCA GTC AGC ACC CTC GAG ACG AGC CCT ATA GTG AGT CGT ATT A-3') were created in an analogous fashion. DNA template (10 μM) and T7 primer (10 μM) were annealed in 10 mM Tris-HCl, pH 8, and 10 mM $MgCl_2$. The DNA was heated to 98°C for 5 minutes and then allowed to cool to RT in the heating block. The annealed template was added to a transcription reaction (0.5 μM final concentration) containing 40 mM Tris-HCl, pH 8, 25 mM $MgCl_2$, 1 mM Spermidine, 5 mM DTT, 0.01% (v/v) triton X-100, 3 mM each NTP, and 16 units of RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, centrifuged to remove insoluable inorganic phosphate accumulated during the reaction, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended into loading dye/water (1:3) and loaded onto 8% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. RNAs were visualized by UV shadowing and excised from the gel. Gel slices were crushed, resuspended in 0.3 M sodium acetate containing 10% (v/v) phenol, and extracted at 37°C overnight. Following extraction, RNAs were filtered through spin columns to remove gel pieces, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended in distilled water, and quantified by UV absorbance at 260 nm.

Dephosphorylation and 5' end labeling. Transcribed RNAs were 5' dephosphorylated using calf intestinal alkaline phosphatase (CIAP, New England Biolabs). CIAP reactions contained 100-500 pmol RNA, 1×NEB3 buffer, 0.5 units RNase Inhibitor (Roche), and 50 units CIAP. Reactions were incubated at 37°C for 30 minutes, made 0.3 M in sodium acetate, extracted with phenol/chloroform/isoamyl alcohol twice, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended in 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 50 μCi γ-[^{32}P]-ATP, and 5 units T4 polynucleotide kinase (Invitrogen). Labeling reactions were incubated at 37°C for 30 minutes, then loaded directly onto 8% (19:1 acrylamide/bisacrylamide) denaturing gels and purified as described above.

2-3.2. Expression and derivatization of U2AF65-based probes.

Cloning: *(The work in the cloning section was performed by A. Reayi and L. Foong).* A panel of cDNAs representing full length and deletion mutants of U2AF65, each with an N-terminal His_6 tag and a Factor Xa site followed by an introduced Cys were prepared by PCR and cloned into the pET expression system using *Eco*RI and *Hin*dIII. PCR products representing the U2AF N-terminal deletion mutants were cloned into pET plasmid vectors (Novagen) carrying kanamycin resistance. The resulting plasmids contained an N-terminal His_6 tag sequence, followed by a Factor Xa proteolysis site and the N-terminal U2AF deletion mutants. The insertion of the engineered U2AF gene was confirmed by restriction digestion of the pET plasmid. Successfully ligated plasmids were transfected into *E.coli* BL21 pLysS cells (Novagen) carrying chloramphenicol resistance. Transformed cells were selected by growth on LB-kan/chl agar plates. Picked colonies were grown in 10 mL LB-kan/chl media overnight at 37°C. Aliquots were made 20% (v/v) with glycerol, flash frozen and stored at -80°C.

Expression of recombinant U2AF65-based probes. Overnight samples 10 mL were grown in LB-kan/chl at 37°C. Overnights were added to 1 L of fresh LB-kan/chl media and grown at 37°C until an absorbance of 0.4-0.6 at 600 nm was reached. Upon reaching log phase of growth, protein expression was induced by the addition of 800 μL 1 M IPTG, followed by incubation at 37°C for an additional 4 hours. Cell cultures were harvested by centrifugation at 12,000 rpm (Beckman, Avanti centrifuge, J-20 rotor) for 30 minutes. The supernatant was discarded and the cell pellets resuspended in 100 mL of lysis buffer (50 mM KH_2PO_4, pH 8, 300 mM KCl, 10 mM imidazole, 0.1% (v/v) tween, 10% (v/v) glycerol, and 10 mg lysozyme). The cell pellets were lysed by stirring on ice for 30 minutes. Following lysis, the mixture was sonicated (Branson Sonifier 450) on ice with 6×10 seconds bursts on 60% output power. The mixture was transferred to a centrifugation tube and centrifuged at 10,000 rpm for 30 minutes. The supernatant was applied to washed Ni-NTA resin (Qiagen) and rotated at 4°C for 60 minutes. Approximately 50 mL of supernatant was added to 6 mL bed volume of Ni-NTA beads. After binding to the resin, the solution was centrifuged at 3000 rpm to pellet the resin and washed with 50 mL wash buffer (50 mM KH_2PO_4, pH 8, 300 mM KCl, 10 mM imidazole, 0.1% (v/v) tween, 10% (v/v) glycerol). The partially cleared solution was transferred to an empty purification column and rewashed with 50 mL of wash buffer. The U2AF was eluted from the column using wash buffer with 250 mM imidazole and collecting 5 mL fractions. The protein containing fractions were pooled and concentrated in Amicon Ultra concentrators (Millipore) down to approximately 10 mL and dialyzed against buffer D or Factor Xa cleavage buffer (3 × 1L). After dialysis, U2AF was flash frozen in 100 μL aliquots and stored at -80°C. U2AF proteins were visualized by running approximately 2 μg of protein on 12% (29:1

acrylamide/bisacrylamide) SDS page gels (375 mM Tris-HCl, pH 8.8, and 0.1% (v/v) SDS).

Factor Xa proteolysis of U2AF65-based probes. U2AF65-based probes were engineered with a Factor Xa proteolysis site at the N-terminus of the protein. Cleavage with Factor Xa exposed an N-terminal cysteine for modification. Factor Xa reactions cleavage reactions contained 3 μg of U2AF65-based probes dialyzed into Factor Xa cleavage buffer (20 mM HEPES, pH 7.9, 1 mM $CaCl_2$, 100 mM KCl, and 20% (v/v) glycerol) and 1/100 to 1/5 ratio (mg) of Factor Xa (Roche) to recombinant protein. Reactions were incubated at room temperature for 3 hours and analyzed by SDS PAGE. SDS loading dye was added to the reactions and fractionated on a 12% (29:1 acrylamide/bisacrylamide) SDS denaturing gel. The optimal Factor Xa to U2AF ratio was determined to be 1/20. Large scale protein digestions contained approximately 3 mg of U2AF65-based probe dialyzed into Factor Xa cleavage buffer and 0.15 μg Factor Xa. Reactions were incubated at room temperature for 3 hours, and then added to 1.5 mL Factor Xa removal resin (Qiagen) and nutated at 4°C for 30 minutes to remove Factor Xa. The resin was pelleted by centrifugation; the supernatant was removed, added to 600 μL of Ni-NTA resin (Qiagen) and nutated at 4°C for 30 minutes. The small cleaved N-terminal fragments and uncleaved protein were removed by incubation with Ni-NTA resin which bound the His_6 tagged proteins. The resin was pelleted by centrifugation, the supernatant removed and dialyzed against buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT, 0.2 mM EDTA, and 20% (v/v) glycerol) for gel shift analysis or buffer D lacking EDTA and glycerol for Fe-EDTA cleavage reactions.

Synthesis of EDTA-3MPA. 3-mercaptopropionic acid (1.0 mL, 11.7 mmol) was added to a stirring suspension of EDTA dianhydride (3.0 grams, 11.7 mmol) in dimethyl formamide (DMF, 15 mL). The reaction was stirred overnight under a nitrogen atmosphere. The reaction was then heated to 60°C and the DMF removed under pressure with a roto-vap. The resulting yellow oil was resuspended in 12 mL 1 M NaOH and stirred at room temperature overnight. The resulting slurry was centrifuged to remove the precipitate, and the supernatant added to 33 mL of methanol and allowed to stand at 4°C for several hours to precipitate EDTA. The reaction was re-centrifuged to remove the precipitate and the supernatant lyophilized to a fine yellow powder. The powder containing EDTA-3MPA was resuspended in distilled water and the pH adjusted to pH 2 with trifluoroacetic acid (TFA). The EDTA-3MPA was purified by C-18 analytical sep-pak (Waters). The sep-pak was pre washed with 50 mL acetonitrile, 20 mL buffer A (2% (v/v) acetonitrile, 0.06% (v/v) TFA) and 20 mL distilled water. The EDTA-3MPA containing solution was added to the column, washed with distilled water (2×50 mL), buffer A (2×50 mL) and eluted from the column with buffer B (90% (v/v) acetonitrile, 0.06% (v/v) TFA). EDTA-3MPA containing fractions were identified by UV absorbance (λ_{max} 236 nm), pooled and lyophilized to a fine powder (1.0 gram, 33% yield). For derivatization reactions, EDTA-3MPA was dissolved in 10% (v/v) DMSO and 100 mM HEPES, pH 8.5 to a final concentration of 50 mM, frozen and stored at -20°C.

Derivatization with EDTA-3MPA. Cleaved U2AF65-based probes in Factor Xa buffer were derivatized in reactions containing 300 mM DTT, and 1 mM EDTA-3MPA or biotin-3MPA. Derivatizations were incubated at 4°C for 12 hours. Biotin-3MPA *(synthesized by A.M.M.)* was dissolved in 100 mM HEPES, pH 8.5 in 5 mM stock solutions. Biotin derivatization reactions were chased with 1 mM EDTA-3MPA and incubated at 4°C for an additional 12 hours. Mock reactions used for cleavage controls were handled the same way as derivatization reactions, however mock reactions contained no EDTA-3MPA. Following derivatization, EDTA modified U2AF proteins were dialyzed into buffer D (20 mM HEPES, pH 7.9, 100 mM KCl, 0.5 mM DTT, 0.2 mM EDTA, and 20% (v/v) glycerol) for gel shift analysis or buffer D lacking EDTA and glycerol for Fe-EDTA cleavage reactions. Factor Xa cleavages and derivatization reactions were monitored by SDS PAGE and western blotting respectively. Approximately 2 μg of protein derivatized with EDTA-3MPA or biotin-3MPA and chased with EDTA-3MPA were fractionated on a 12% (29:1 acrylamide/bisacrylamide) SDS PAGE gel. The gel was then soaked in transfer buffer (25 mM glycine, 40 mM monoethanolamine, and 20% (v/v) methanol). The gel was sandwiched between soaked filter paper and nitrocellulose transfer membrane (Osmonics), stirred at 4°C, 200 mA for

3 hours. The SDS PAGE gel was stained with Coomassie blue to visualize all proteins. The nitrocellulose was rinsed with wash buffer (20 mM Tris-HCl, pH 7.6, 140 mM NaCl, and 0.1% (v/v) Tween-20). The nitrocellulose was blocked by incubation in blocking buffer (0.05% (w/v) milk powder, 0.2% (v/v) Tween-20, 20 mM Tris-HCl, pH 7.6, and 140 mM NaCl) and rotated at 4°C for 1 hour. Following blocking, the nitrocellulose was probed with 50 μL avidin-HRP-conjugate (BIO-RAD) in modified blocking buffer (0.02% (w/v) milk powder, 0.2% (v/v) Tween-20, 20 mM Tris-HCl, pH 7.6, and 140 mM NaCl) and further rotated at 4°C for 1 hour. The nitrocellulose was washed twice with wash buffer and then developed with a pre-mixed developing solution (5.6 mM 4-chloro-1-napthol, 20% (v/v) methanol, 5 μM H_2O_2, 20 mM Tris-HCl, pH 7.6, and 140 mM NaCl). The gel was allowed to develop until the protein bands were the correct intensity and then washed with distilled water.

2-3.3. Binding and affinity cleavage.

U2AF gel shift experiments. Binding reactions containing 5' ^{32}P-labeled RNA (50-100 × 10^3 c.p.m.) and 0-100 pmol derivatized U2AF65 were performed in 10 mM HEPES, pH 7.9, 60 mM KCl, 2 mM $MgCl_2$, 1 μg tRNA, 0.25 mM DTT, 0.1 mM EDTA and 10% (v/v) glycerol. Reactions were incubated at room temperature for 60 minutes and then immediately analyzed by 6% native PAGE (89:1 acrylamide/bis-acrylamide) run with 50 mM Tris-glycine running buffer at 110V for 3 hours. Multi-protein–RNA complexes containing U2AF65, U2AF35 and/or SF1-C4 were analyzed in an analogous fashion. Recombinant His_6 tagged U2AF35 and SF1-C4 mutant were over-expressed in *E. coli* and purified by Ni-NTA chromatography as described above for U2AF65-based probes.

RNase protection/footprinting. Assays were performed in binding reactions prepared as described above with 1.0 μM U2AF followed by addition of 0.5 ng RNase A (Roche) or 0.3 units RNase T1 (Roche) and digested at room temperature for 3 minutes and 1 minute respectively. Reactions were quenched with phenol/chloroform/isoamyl alcohol, extracted and ethanol precipitated. Following ethanol precipitation, reactions were subjected to 10% denaturing PAGE (19:1 acrylamide/bis-acrylamide). Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Affinity cleavage reactions. Reactions contained 0.6 μM U2AF-EDTA (dialyzed into buffer D lacking EDTA and glycerol) in binding reactions as described above. Binding was performed at room temperature for 60 minutes followed by the addition of 0.5 μM $FeSO_4$ and a further 10 minute incubation on ice to allow chelation. Affinity cleavage was initiated by the addition of 0.05% (v/v) H_2O_2 and 5 mM ascorbic acid and the reactions were allowed to proceed on ice for 10 minutes. Reactions were quenched with 30 mM thiourea containing 1% (v/v) glycerol, extracted with phenol/chloroform/isoamyl alcohol, and ethanol precipitated. Following ethanol precipitation, reactions were subjected to 10% denaturing PAGE (19:1 acrylamide/bis-acrylamide). Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager. For cleavage experiments in the context of multi-protein complexes, His_6 tagged U2AF35 and SF1-C4 mutant were dialyzed into buffer D lacking EDTA and glycerol.

2-4. *References.*

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-412.

Banerjee, H., Rahn, A., Davis, W., and Singh, R. (2003). Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* **9**, 88-99.

Berglund, J.A, Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.

Black, D.L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103**, 367-370.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). In The RNA World 2nd Ed. (eds Gesteland, R.F., Cech, T.R., and Atkins, J.F.) 525-560 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Chan, R.C., and Black, D.L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.* **17**, 4667-4676.

Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888-896.

Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779.

Erlanson, D.A., Chytil, M., and Verdine, G.L. (1996). The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex. *Chem. Biol.* **3**, 981-991.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.

Ito, T., Muto, Y., Green, M.R., and Yokoyama, S. (1999). Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)). *EMBO J.* **18**, 4523-4534.

Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.* **9**, 576-581.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**, 595-605.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.

Lee, C.G., Zamore, P.D., Green, M.R., and Hurwitz, J. (1993). RNA annealing activity is intrinsically associated with U2AF. *J. Biol. Chem.* **268**, 1342-1348.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.

Maniatis, T., and Tasic, B., (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

Query, C.C., McCaw, P.S., and Sharp, P.A. (1997). A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. *Mol. Cell. Biol.* **17**, 2944-2953.

Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P., and Krämer, A. (1998). Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *RNA* **4**, 551-565.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113-2123.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**, 340-345.

Rudner, D.Z., Breger, K.S., Kanaar, R., Adams, M.D., and Rio, D.C. (1998). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. *Mol. Cell. Biol.* **18**, 4004-4011.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF(65) and SF1/mBBP. *Mol. Cell* **11**, 965-976.

Singh, R., Valcarcel, J., and Green, M.R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173-1176.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**, 7670-7682.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706-1709.

Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171-175.

Wang, X., and Tanaka Hall, T.M. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **8**, 141-145.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

Wu, S., Romfo, C. M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-838.

Chapter 3.

Characterization of a U2AF independent commitment complex (E') in the mammalian spliceosome assembly pathway.

Chapter 3. Characterization of a U2AF independent commitment complex (E') in the mammalian spliceosome assembly pathway.

3-1. Introduction.

3-1.1. The commitment complex.

Early recognition of pre-mRNA substrates by the splicing machinery proceeds through formation of the ATP independent commitment complex (CC) in yeast or early (E) complex in mammals (Legrain et al., 1988; Michaud and Reed, 1991). In both yeast and mammals the 5' splice site is initially recognized by U1 snRNP; the branch region and 3' splice site are defined by the association of the branch binding protein (BBP) and Mud2 (yeast) or SF1 and the heterodimer U2AF (mammals). In the presence of ATP, U2 snRNP becomes stably associated with the pre-mRNA through a base-pairing interaction between U2 snRNA and the branch region; subsequent recruitment of the U4/U6/U5 tri-snRNP and several snRNA rearrangements result in the formation of the mature spliceosome (Staley and Guthrie, 1998).

The mammalian E complex can be visualized either by gel filtration or native gel electrophoresis (Das and Reed, 1999; Michaud and Reed, 1993). In yeast, two commitment complexes, CC1 and CC2 have been identified in U2 snRNP depleted extracts by gel electrophoresis (Legrain et al., 1988; Serephin and Rosbash, 1989, Zhang and Rosbash, 1999). Formation of the faster mobility complex CC1 is dependent on the presence of a 5' splice site and U1 snRNP. The lower mobility complex CC2 is dependent on the presence of a 5' splice site as well as a branch point sequence. Several components of CC2 have been identified. These include the splicing factor Mud2p that has been shown to interact with the highly conserved yeast branch point sequence and U1 snRNP (Abovich and Rosbash, 1997; Zhang and Rosbash, 1999). Mud2p has been identified as the yeast homolog of U2AF65, the large subunit of U2AF, a component of E complex that interacts with the polypyrimidine tract 3' to the highly degenerate mammalian branch point sequence (Zamore et al., 1992). The yeast BBP has been found to directly interact with Mud2p, which parallels the association of the mammalian splicing factor SF1 with U2AF65 (Berglund et al., 1998; Rutz and Serephin, 1999). Taken together, these results suggest a similar structural organization of the yeast and mammalian commitment complexes.

Previously, it has been demonstrated that initial recognition of the pre-mRNA and formation of the commitment complex in both the yeast and mammalian systems is dependent on U1 snRNP. Depletion experiments have demonstrated that the stable association of U2 snRNP with the pre-mRNA requires the presence of U1 snRNP (Barabino et al., 1990). There is good evidence to suggest U1 snRNP is involved in recruiting U2AF to the polypyrimidine tract, which in turn promotes the association of U2 snRNP with the branch region (Cote et al., 1995; Li and Blencowe, 1999). For example, the 5' splice site of exon 17 of the pre-mRNA for the neural cell adhesion molecule has been found to influence U2AF65 binding the 3' splice site of exon 19 (Cote et al., 1995). These results demonstrated a U2AF65/U1 snRNP interaction that occurs at the commitment stage of spliceosome assembly and plays a role in exon 18 skipping. Furthermore, binding of U2AF65 to the polypyrimidine tract of an exonic splicing enhancer (ESE)-dependent pre-mRNA is dependent on the presence of U1 snRNP (Li and Blencowe, 1999). This result indicates that U1 snRNP functions in stabilizing the binding of U2AF65 to the pre-mRNA polypyrimidine tract. The stable association of U1 snRNP with the 5' splice site involves base-pairing interactions between the pre-mRNA and the 5' end of U1 snRNA. Recently, it has been suggested that in yeast this RNA•RNA interaction is preceded by the sequence specific recognition of the 5' splice site by the U1 snRNP protein U1C (Du and Rosbash, 2002).

Here, we report the identification of a novel pre-spliceosomal complex designated E' which forms in U2AF depleted HeLa nuclear extracts. The E' complex is dependent on U1 snRNA-5' splice site base-pairing and commits the pre-mRNA to the splicing pathway. We have probed the structure of complexes formed in wild type and U2AF depleted extracts using RNAs site-specifically derivatized with the hydroxyl radical source Fe-BABE. These experiments indicate that the close proximation of the 5' splice site and branch region observed in E complex are also found in the E' complex and support a model whereby U1 snRNP directly recruits U2AF to the pre-mRNA substrate.

3-2. Results and Discussion.

3-2.1. Identification of a novel pre-spliceosomal complex.

Work by Reed and co-workers has demonstrated that a complex formed on pre-mRNAs corresponding to the previously characterized early or E complex can be resolved from the non-specific H complex by native agarose gel electrophoresis (Figure 3-1A, lane 2; Das and Reed, 1999; Michaud and Reed, 1993). As part of a series of investigations of the earliest steps of spliceosome assembly, we specifically depleted HeLa nuclear extracts of U2AF (MacMillan et al., 1997; Zamore et al., 1992) and analyzed complex formation on the PIP85.B pre-mRNA by native agarose gel electrophoresis. Following incubation of pre-mRNA in depleted extract in the absence of ATP for 30 minutes at 30°C, we were able to observe formation of a complex of slower mobility than H, distinct from E, which we designated E' (Figure 3-1A, lane 6). Formation of both the E and the E' complexes involves U1 snRNA-5' splice site base-pairing since pre-incubation of extract with an anti-sense 2'-O-methyl oligonucleotide complementary to the 5' end of U1 snRNA abolished complex formation (Figure 3-1A; lanes 3 and 7). Similarly, pre-treatment of extract with RNase H and a DNA oligonucleotide directed against the 5' end of U1 snRNA abrogated the formation of both E and E' complexes (see Chapter 4). Blocking the interaction of U2 snRNP with the pre-mRNA using an anti-sense oligonucleotide had no effect on either E or E' formation (Figure 3-1A, lanes 4 and 8). Furthermore, E' complex was also detected on pre-mRNA substrates that contained a mutant BPS (Figure 3-1B; lane 3). This result is consistent with the observation that E complex forms on mutant BPS pre-mRNAs (Champion-Arnaud et al., 1995).

In order to determine which pre-mRNA sequence elements were required for E' complex formation, mutant RNAs were prepared in which one or more sequence elements were eliminated. We challenged E' complex formation on wild type body-labeled pre-mRNA in U2AF depleted nuclear extract pre-incubated with a 500 fold excess of cold mutant RNAs (Figure 3-1C).

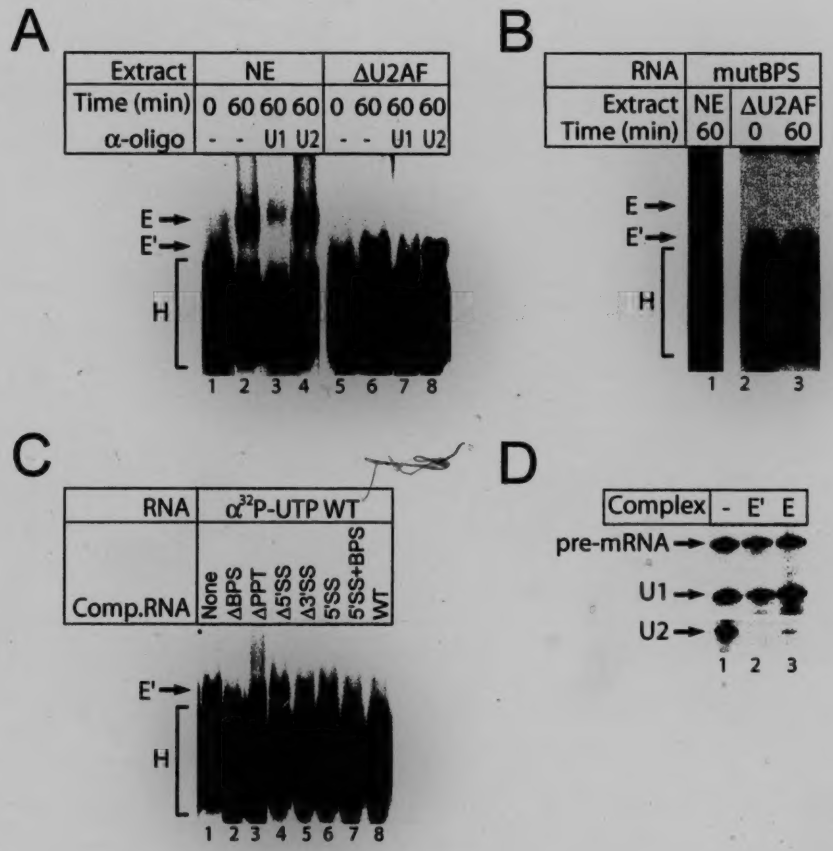


Figure 3-1. Identification of a novel pre-spliceosome complex (E') that contains U1 snRNP. **(A)** Native agarose gel shift analysis of complexes formed in nuclear extract (NE) and U2AF depleted (ΔU2AF) nuclear extract. Complex formation was analyzed in the presence and absence of anti-U1 and anti-U2 oligonucleotides. **(B)** E' complex formation on mutBPS pre-mRNAs. Native agarose gel shift analysis of complexes formed in nuclear extract (NE) and U2AF depleted (ΔU2AF) nuclear extract. **(C)** E' formation is dependent on the presence of a 5' splice site within the pre-mRNA. Complexes were formed on a body-labeled wild type pre-mRNA (α^{32}P-UTP WT) in extract saturated with unlabeled mutant competitor RNA. **(D)** E' complex contains U1 snRNP. RT primer extension of RNAs purified from isolated complexes. (-) RNA isolated from an aliquot of precipitated nuclear extract. The pre-mRNA, U1 snRNA, and U2 snRNA are indicated.

Competitor RNAs containing a 5' splice site inhibited E' formation but no inhibition was observed using substrates that lacked this sequence element (Δ5'SS) indicating that a 5' splice site is both necessary and sufficient for formation of the E' complex. In order to confirm that the E' complex contains U1 snRNP, both E' and E complexes were purified from agarose gels and the isolated RNA subjected to reverse transcription with the appropriate primers. U1 snRNA was found to be present in both E' and E complex (Figure 3-1D), indicating that E' complex contains U1 snRNP; we were unable to detect U2 snRNA in either complex (Figure 3-1D).

3-2.2. Relationship of E' to the pre-spliceosome.

The formation of E complex commits the bound pre-mRNA to the splicing pathway. Thus, prior formation of E complex at 30°C on a ^{32}P labeled substrate results in efficient splicing of the labeled RNA even after the addition of 500-fold excess of competitor RNA (Legrain et al., 1988).

We tested the commitment ability of both wild-type and U2AF depleted extracts by pre-incubation of pre-mRNA in extract followed by a chase with extract containing an excess of competitor RNA. Direct incubation of a radio-labeled pre-mRNA substrate in extract in the presence of increasing concentrations of unlabeled competitor RNA resulted in the decrease of spliced products (Figure 3-2A; lanes 1-4). When E complex was formed by pre-incubation of substrate RNA in wild type extract, followed by addition of an extract mix containing competitor, no inhibition of splicing was observed (Figure 3-2A; lanes 5-8). Commitment correlated with E complex formation since pre-incubation of a pre-mRNA in wild-type or U2AF depleted extract on ice followed by a chase with wild-type extract mix containing competitor RNA resulted in significantly decreased splicing (Figure 3-2A; lanes 9-12, and lanes 13-16). We tested whether the formation of E' commits pre-mRNA to the splicing pathway by pre-incubation of labeled substrate RNA in U2AF depleted extract under E' forming conditions followed by a chase with wild-type extract containing competitor (Figure 3-2A; lanes 17-20). Under these conditions, normal splicing was observed indicating that E' is a functional commitment complex.

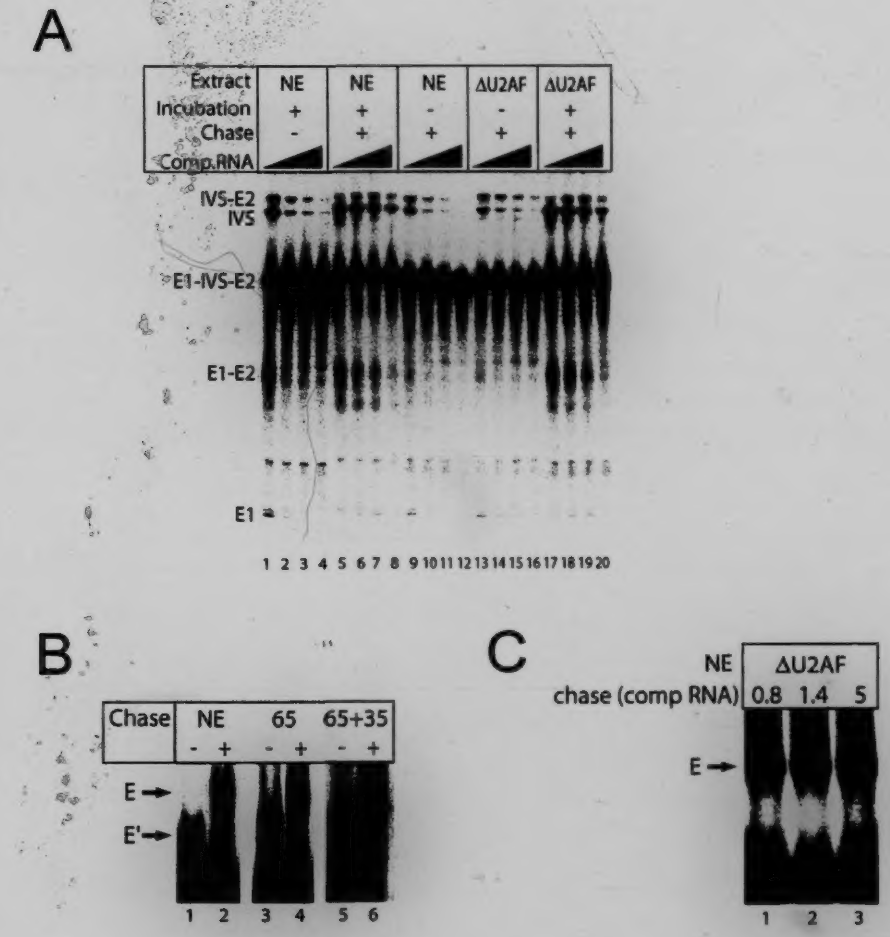


Figure 3-2. Commitment of pre-mRNA to splicing in the absence of U2AF. **(A)** E' complex contains commitment ability. RNAs were preincubated in nuclear extract (NE), or U2AF depleted (ΔU2AF) nuclear extract at 30°C (+) or 0°C (-) as indicated, and chased with wild type nuclear extract (+) saturated with increasing competitor RNA (0, 0.8, 1.4, 5 μM). Splicing substrates, products, and intermediates are indicated. **(B)** E' complex is a precursor to E complex. Pre-formed E' complex chased with nuclear extract (NE, lanes 1 and 2), U2AF65 (lanes 3 and 4) or U2AF65/35 heterodimer (lanes 5 and 6) shifts E' complex to E complex. **(C)** U2AF is recruited to E' complex in the presence of competitor RNA. Pre-formed E' complex is shifted to E complex with wild type NE chase in the presence of excess competitor RNA. Competitor RNA (0.8, 1.4, 5 μM) added to a nuclear extract chase mix was added to pre-formed E' complex and the reactions analyzed by native agarose gel electrophoresis.

We investigated the relationship between E' and E complexes by pre-forming E' on a substrate pre-mRNA, chasing with either wild-type extract or recombinant U2AF and then analyzing the complexes by native agarose gel electrophoresis (Figure 3-2B). Following the formation of E' complex, addition of wild type nuclear extract efficiently chased E' into E complex (Figure 3-2B; lane 2). Furthermore, addition of U2AF65 alone or recombinant U2AF, efficiently chased E' into E (Figure 3-2B; lanes 4 and 6) suggesting that E' is a functional precursor to E in the commitment of pre-mRNA to the splicing pathway. Furthermore, after the formation of E' complex, incubation with wild type nuclear extract saturated with cold competitor RNA had no effect on the conversion of E' to E complex (Figure 3-2C). This data is consistent with the observation that E' complex is able to commit pre-mRNA to splicing.

3-2.3. Directed hydroxyl radical probing of pre-mRNA structure in E'.

We probed the structure of RNA bound in the E' complex using pre-mRNA site-specifically derivatized with a directed hydroxyl radical probe, Fe-BABE, tethered to the pre-mRNA (Chapter 1; Kent and MacMillan, 2002). Using a wild type RNA construct that had been derivatized with Fe-BABE at either the branch region (WT148) or 3' splice site (WT179) and a mutant BPS pre-mRNA derivatized at the branch region (148) we probed complexes formed in wild-type and U2AF depleted extracts (Figures 3-3A, 3-3B and 3-3C). Reactions were incubated under E' or E complex formation conditions (lacking ATP and $MgCl_2$) followed by Fe-BABE mediated cleavage and analysis of cleavage products by RT primer extension. As previously reported, under E complex forming conditions, probes tethered to either the branch region or 3' splice site generated strong cleavages at the 5' splice site (Figure 3-3B; lanes 5 and 8) indicating the close proximation of these sequence elements in E complex. No cleavages at the 5' splice site were observed from either WT148 or WT179 derivatized pre-mRNA when incubated in nuclear extract on ice (Figure 3-3B; lanes 4 and 7), indicating the observed cleavages are due to the formation of E complex. Interestingly, strong 5' splice site cleavages were observed under E' forming conditions in U2AF depleted extract using pre-mRNAs derivatized at the branch region (Figure 3-3B; lane 6).

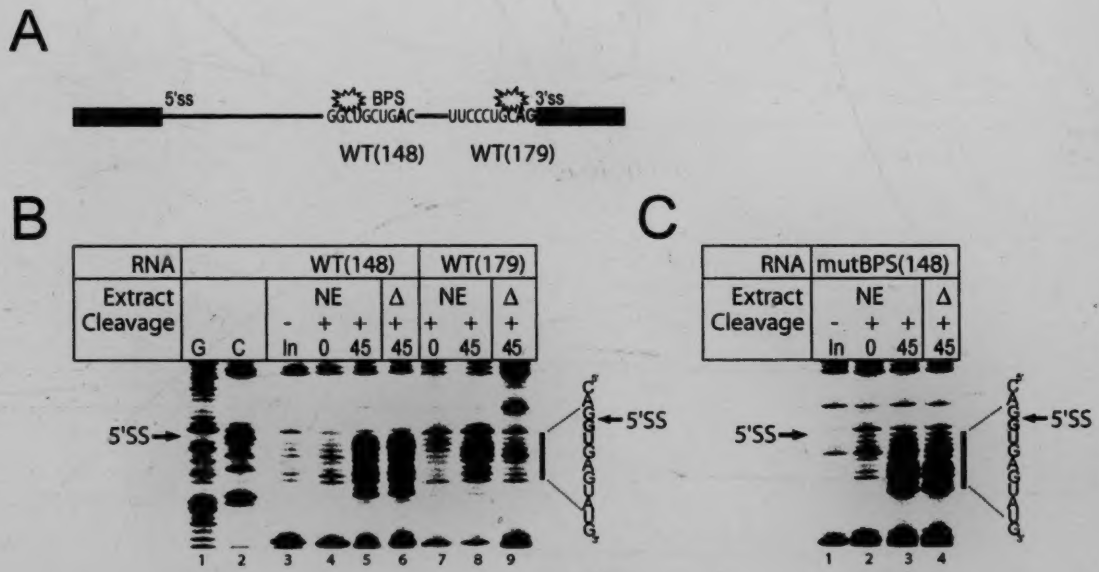


Figure 3-3. Directed hydroxyl radical cleavage of ATP independent pre-spliceosome complexes. **(A)** pre-mRNA substrate derivatized with Fe-BABE (star) at the branch region (WT148) or 3' splice site (WT 179). **(B)** Analysis of cleavage reactions performed in nuclear extract (NE) and U2AF depleted (Δ) nuclear extract probed with WT148 and WT179 pre-mRNAs. Reactions were probed after incubation at 30°C for 0 or 45 minutes. RT reactions were compared to input RNA (In) for location of RT stops. G and C sequencing lanes shown. Sequence around the 5' splice site indicated. Regions of significant cleavage represented by the vertical bar. **(C)** Analysis of mutBPS pre-mRNAs derivatized at position 148 with Fe-BABE. Cleavage reactions performed in nuclear extract (NE) and U2AF depleted (Δ) nuclear extract probed under E or E' complex conditions. Reactions were probed after incubation at 30°C for 0 or 45 minutes. Sequence around the 5' splice site indicated. Regions of significant cleavage represented by the vertical bar.

In contrast, no 5' splice site cleavage was observed under the same conditions using the probe derivatized at the 3' splice site (Figure 3-3B; lane 9). These results suggest that the branch region may be held in proximity to the 5' splice site through an interaction between SF1 and U1 snRNP.

The mutant BPS pre-mRNA derivatized at position 148 with Fe-BABE probed in E complex generated strong cleavages at the 5' splice site as reported previously (Figure 3-3C; lane 3; Chapter 1; Kent and MacMillan, 2002). Cleavage at the 5' splice site was not observed when the substrate was incubated on ice, indicating the observed cleavages are due to the formation of the commitment complex (Figure 3-3C; lane 2). Strong cleavages at the 5' splice site were also observed when the mutant BPS substrate was incubated in U2AF depleted nuclear extract and probed under E' complex conditions (Figure 3-3C; lane 4). This suggests a similar structural organization of the branch region and 5' splice site in both E and E' complexes.

3-2.4. Detection of branch binding protein in E' complex.

Since pre-mRNAs containing a mutant BPS can assemble into both E and E' complex, and the branch region is held in proximity to the 5' splice site within the E' complex, we wanted to determine if SF1 was associating with the pre-mRNA branch region in the absence of U2AF. We created an RNA construct containing a reactive thiol at the BPS (Chapter 1; Appendix II; Kent and MacMillan, 2002) with a ^{32}P label facilitating analysis by RNase A digestion (Figure 3-4A). The thiol was derivatized with the efficient photo crosslinker benzophenone maleimide, (MacMillan et al., 1994). E' and E complexes were formed on derivatized pre-mRNAs and UV photolyzed to crosslink proteins proximal to the branch region. Following RNase A digestion, specific RNA-protein crosslinks obtained from purified complexes were analyzed by SDS PAGE (Figure 3-4B). In E complex three proteins can be efficiently crosslinked to the branch region. Based on molecular weight, these proteins likely correspond to the branch binding protein SF1, U2AF65, and U2AF35 respectively (Figure 3-4B; lane 3). We also detected a minor crosslinked protein of approximately 90 kDa apparent molecular weight.

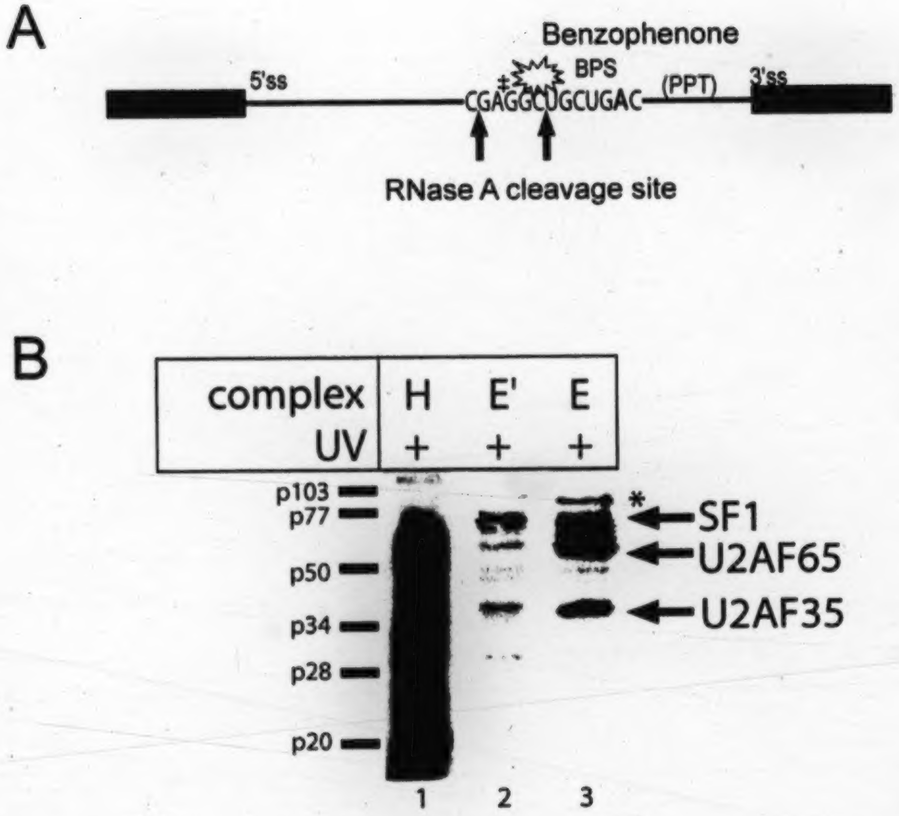


Figure 3-4. Site-specific protein crosslinking at the branch region. **(A)** A pre-mRNA was constructed to contain benzophenone (star) site specifically incorporated in the branch point region. The RNA contains a ^{32}P label (‡) between RNase A cleavage sites (arrows). **(B)** Specific protein-RNA crosslinking. Derivatized RNAs were assembled into H (lane 1), E' (lane 2) and E (lane 3) complexes, irradiated with UV and isolated from agarose gels. Samples were digested with RNase A and analyzed by SDS PAGE. Indicated based on molecular weight are SF1, U2AF65 and U2AF35. Molecular weight markers indicated. Asterisk represents a ~90kDa spliceosome associated protein (SAP; Bennett et al., 1992; Burns et al., 1999; Zhou et al., 2002).

This protein may represent CDC5, a protein associated with pre-spliceosomes (Bennett et al., 1992; Zhou et al., 2002) that has been implicated in the first chemical step of splicing (Burns et al., 1999). In the E' complex, only the branch binding protein SF1 can be detected (Figure 3-4B; lane 2). This suggests that E' complex contains SF1 bound to the branch region in the absence of U2AF.

3-2.5. E' complex is the mammalian equivalent to the yeast CC1.

We have detected the formation of a novel commitment complex, designated E', on pre-mRNA substrates incubated in U2AF depleted extracts. The E' complex forms in the absence of ATP and contains U1 snRNP suggesting that it is analogous to the yeast commitment complex CC1. Our results demonstrate that E' complex may be similar to CC1 in several ways: formation of both E' complex and CC1 requires a functional 5' splice site and the presence of U1 snRNP. The formation of E' complex does not require a functional branch sequence mirroring the requirement for formation of CC1. Furthermore, since the overall pathways of spliceosome assembly in yeast and mammals are highly conserved (Abovich and Rosbash, 1997), it is likely that formation of the E complex and the yeast commitment complex follow a similar assembly pathway. The observation of E' in HeLa extracts suggests a further parallel with the yeast splicing system: the assembly of the mammalian spliceosome may proceed through the formation of two ATP independent commitment complexes.

Experiments in the yeast system have demonstrated that the commitment of pre-mRNAs to splicing is dependent on U1 snRNP and that the complexes formed in U2 snRNP depleted extracts are intermediates in spliceosome assembly (Legrain et al., 1988; Serephin and Rosbash, 1989). Similarly, our results suggest that commitment of mammalian pre-mRNAs to splicing is dependent upon the initial interaction of U1 snRNP with the 5' splice site and the interaction of SF1 at the branch region followed by the recruitment of U2AF to the polypyrimidine tract and 3' splice site to give E complex (Figure 3-5).

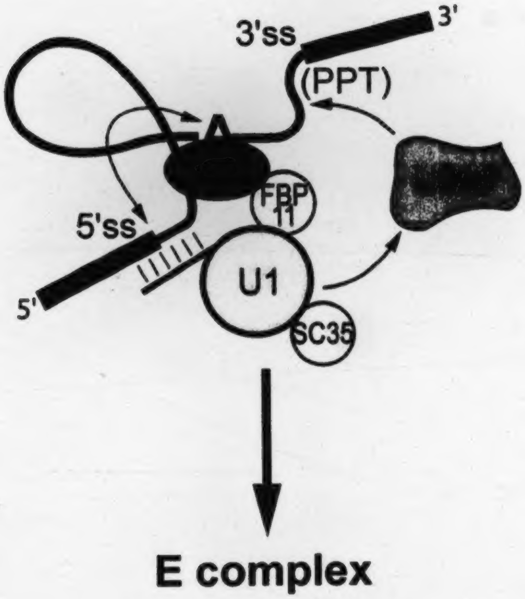


Figure 3-5. E complex assembly. Assembly proceeds through early recognition of the 5' splice site followed by recruitment of U2AF to the polypyrimidine tract. Pre-mRNAs are committed to pre-spliceosome assembly by the association of U1 snRNP with the 5' splice site. SF1 interacts with the branch point region and is bridged to U1 snRNP via an interaction with FBP11 which results in the close proximity of the BPS and 5' splice site (green arrow). Stable U1 snRNA base pairing with the 5' splice site facilitates the formation of E' complex. U1 snRNP and/or SF1 recruit U2AF to the polypyrimidine tract resulting in E complex formation.

Since many protein•protein interactions have been observed in the commitment complex (Abovich and Rosbash, 1997; Berglund et al., 1998; Reed, 2000), it is possible that U2AF is recruited to the E' complex through interactions with both SF1 and U1 snRNP.

Although no sequences 3' to the BPS in yeast have been shown to be essential for commitment complex formation, the formation of the mammalian E complex is dependent on the presence of a functional polypyrimidine tract. It has been reported that yeast commitment complex formation (CC2) can be detected in ΔMud2p extracts, but not in extracts lacking BBP (Abovich and Rosbash, 1997; Abovich et al., 1994). Our results demonstrate the detection of the E' complex in extracts depleted of U2AF. These observations complement each other given that yeast splicing is dependent on a functional branch sequence and mammalian splicing is dependent on a functional polypyrimidine tract.

The precise determinants of commitment in the pre-spliceosome are not fully understood. The studies outlined here indicate that commitment occurs in the absence of U2AF but confirm that commitment ability of E' requires U1 snRNP association with the 5' splice site. Although the U1 snRNP•pre-mRNA association in E' requires base-pairing between the U1 snRNA and the 5' splice-site, actual commitment probably precedes this association. For example, a pre-mRNA incubated in extract where the 5' end of U1 snRNA has been removed by RNase H digestion does not splice when added to a reaction containing wild-type U1 snRNP suggesting "commitment" in the digested extract (see Chapter 4).

3-2.6. Structural organization of the E' complex.

We have previously observed a close proximation of both the branch region and 3' splice site with the 5' splice site in E complex using pre-mRNA probes modified with Fe-BABE (Kent and MacMillan, 2002). Similar experiments, under E'-forming conditions, in the absence of U2AF, indicate that the 5' splice and branch region are close to one another in E' and probably reflect bridging interactions between U1 snRNP and the branch binding protein SF1 (Figure 3-5). The interactions between SF1, the mammalian Prp40-like protein FBP11, and U1 snRNP have been proposed to bridge the 5' splice site and the branch region in E complex (Abovich and Rosbash, 1997; Reed, 2000). The fact that no 5' splice-site cleavages are observed in E' using a pre-mRNA probe derivatized at the 3' splice site is consistent with the role of U2AF as a 3' splice site recognition factor which brings that portion of the RNA into direct proximity with the branch region and hence 5' splice site (Kent et al., 2003). Previously, it has been demonstrated that U2AF35 interacts with the U1 snRNP specific protein U1-70K via an interaction with the SR protein SC35, which serves to bridge the two ends of the intron (Wu and Maniatis, 1993). U1 snRNP association with the pre-mRNA defines the 5' splice site while the interaction of U2AF65 with the polypyrimidine tract, and specifically, the direct recognition of the AG dinucleotide by U2AF35 define the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zhang and Rosbash, 1999). The proximation of 5' splice site and branch region in the E' complex are consistent with a U1 snRNP facilitated recruitment of U2AF to the pre-mRNA (Cote et al., 1995; Li and Blencowe, 1999) and argue for a very early definition of the intron that precedes this association.

3-3. Materials and Methods.

3-3.1. Synthesis of pre-mRNA substrates.

Plasmid preparation. Competent *E.coli* BL21 cells were transformed with the PIP85.B plasmid containing the PIP85.B pre-mRNA and an ampicillin resistance cassette which was purified as described previously (Chapter 1).

PCR of transcription templates. Templates for transcription were made by PCR using the PIP85.B plasmid. The primers M13F and CQ27 were used to create a PCR product containing a T7 RNA polymerase promoter representing a transcription template for the full length PIP85.B pre-mRNA. Mutant RNAs were transcribed from PCR products based on PCR from the PIP85.B plasmid with primers M13F and CQ36: 5' splice site RNA, CQ5 and CQ27: polypyrimidine tract/3' splice site RNA, CQ32 and CQ27: polypyrimidine tract knock-out/3' splice site RNA. Plasmid and primers were incubated in reactions that contained 10 mM Tris-HCl, pH 9, 1.5 mM $MgCl_2$, 50 mM KCl, 1 mM of each dNTP and 2 units TAQ DNA polymerase as described previously (Chapter 1).

Transcription. RNAs were transcribed from PCR products derived from the PIP85.B plasmid described above. Full length pre-mRNA (13/27), 5' piece (13/36) and 3' piece (5/27) were transcribed from the indicated PCR products. Standard transcription reactions contained 40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM each ATP, CTP, and GTPs, 15 µM UTP, 100 µCi α-[^{32}P]-UTP (3000 Ci/mmol, 800 µCi/uL, NEN), 1 mM DTT, 100 ng PCR derived template, and 2 units RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, quenched with loading dye (0.4 mg/mL Bromo Phenol Blue/Xylene cyanole, 6.4 M urea, 0.4 M Tris-base, 0.4 M boric acid, and 100 mM Na_2EDTA) and loaded onto 8% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. RNAs were visualized by autoradiography using Kodac X-OMAT film, and excised from the gel. Gel slices were crushed, resuspended into 0.3 M sodium acetate containing 10% (v/v) phenol, and extracted at 37°C overnight. Following extraction, RNAs were filtered through spin columns to remove gel pieces, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated with 2 µg glycogen. Precipitated RNAs were resuspended into distilled water and stored at -80°C.

Site specifically modified pre-mRNAs. RNAs that contained the N4-functionalized cytosine (ClΦU) were prepared as previously described (Chapter 1 and Appendix *II*). Oligonucleotides representing the PIP85.B pre-mRNA branch point sequence (5'-GG(ClϕU)UGCUGAC-3'), and representing the PIP85.B 3' splice site (5'-UUCCCUG(ClϕU)AG-3') were incorporated into full length pre-mRNAs using a three piece ligation procedure (Chapter 1; Query et al., 1994) with 5' and 3' pieces of RNA and a DNA bridging oligonucleotide as described previously (Chapter 1).

3-3.2. Analysis of pre-spliceosome complexes and pre-mRNA splicing.

Pre-spliceosome complex formation. For analysis of E complex and E' complex formation, pre-mRNAs (50-100×10^3 c.p.m.) were incubated in reactions containing 25% (v/v) HeLa nuclear extract or 25% (v/v) U2AF depleted HeLa nuclear extract respectively, 60 mM KCl, and 4 units RNase inhibitor. Reactions were incubated at 30°C for 45 to 120 minutes and stopped by the addition of 2 μL 97% (v/v) glycerol loading dye (0.4 mg/mL BPB/XC, 50 mM Tris-Glycine pH 8.6) and incubated on ice. Five μL aliquots were fractionated on 1.2% (w/v) agarose, 0.5×TBE gels. Gels were poured as previously described, dried, and exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager. Specific interactions of U1 snRNP and U2 snRNP with the pre-mRNA were blocked with anti-U1 and anti-U2 biotinylated-2'-O-methyl anti-sense oligonucleotides (Dharmacon) as described elsewhere (Chapter 1; Blencowe and Lamond, 1999). Blocking reactions were performed with anti-sense oligonucleotides added to a final concentration of 3 μM to aliquots of nuclear extract and incubated at 30°C for 30 minutes. Pre-mRNAs were added to blocked nuclear extract and analyzed for complex formation as described above. Add-back experiments involving U2AF were performed in U2AF depleted nuclear extract under E' forming conditions as described above in the presence of 30-60 μM recombinant His_6 tagged U2AF65 and U2AF35 (expressed and purified as described in Chapter 2) or an aliquot of wild type nuclear extract added to pre-formed E' complex. Complex formation assayed in the presence of cold competitor RNAs (0.8-5 μM final concentration) were performed in complex forming reactions as described above with nuclear extracts that had been pre-saturated with the competitor RNA.

Isolation of pre-spliceosome complexes. The E and E' complexes were formed on ^{32}P-body labeled pre-mRNAs and fractionated on agarose gels as described above. Gels were visualized by autoradiography, complexes were excised from the gel, and the gel slices crushed and soaked in 0.3 M sodium acetate at 37°C overnight. Following extraction, isolated RNAs were filtered through spin columns to remove gel pieces, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. The precipitated RNAs were resuspended in distilled water, and subjected to RT primer extension for the presence of pre-mRNA, U1 snRNA and U2 snRNA as describe below.

Preparation of U2AF depleted HeLa nuclear extract. An aliquot of HeLa nuclear extract was dialyzed into 1 M KCl buffer D (20 mM HEPES, pH 7.9, 1M KCl, 0.5 mM DTT, 0.2 mM EDTA, and 10% (v/v) glycerol) with 3000 MWCO dialysis tubing (Fisher Brand) presoaked in water. Dialysis was performed at 4°C for 2×1 L buffer changes and then left overnight stirring at 4°C. Poly (U) sepharose resin (Amersham Phamacia Biotech) was presoaked in 1 M KCl buffer D and nutated overnight at 4°C. The resin was then loaded onto a purification column as a slurry. The resin was allowed to settle and then washed with 50 mL of 1 M KCl buffer D. The high salt nuclear extract was then loaded onto the column and fractions collected. Protein concentration was determined for isolated fractions using a Bradford assay (BIO-RAD). High protein fractions representing the U2AF depleted extract were pooled and dialyzed into 100 mM KCl buffer D. Aliquots were flash frozen and stored at -80°C.

Splicing assays. Pre-mRNAs ($50\text{-}100\times10^3$ c.p.m.) were incubated in 20 μL reactions containing 40% (v/v) HeLa nuclear extract, 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and 4 units ribonuclease inhibitor. Competition assays were performed in the presence of 0.8-5 pmol of competitor RNA. For competition assays, an aliquot of nuclear extract was incubated with competitor RNA at 30°C for 5 minutes and then combined in a standard splicing reaction with body labeled pre-mRNA substrate. Commitment assays were performed by incubating pre-mRNA ($50\text{-}100\times10^3$ c.p.m.) under E or E' complex conditions as described above. Aliquots were then added to a splicing mix containing ATP/CP and $MgCl_2$ with or without competitor RNA. Following incubation at 30°C, splicing reactions were digested with 40 μg proteinase K in the presence of 20 μg tRNA at 55°C for 20 minutes as described previously (Chapter 1). Reactions were extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated, and then subjected to 15% (19:1 acrylamide/bisacrylamide) denaturing polyacrylamide gel electrophoresis. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

3-3.3. Affinity cleavage and UV crosslinking.

Derivatization of N4-functionalized cytosine. Pre-mRNA ligation products containing the N4-functionalized cytosine at the branch region or 3' splice site were derivatized with Fe-BABE (iron mediated RNA cleavage) or benzophenone maleimide (UV crosslinking) under reducing conditions. Ethanol precipitated RNA was resuspended in 2 mM DTT, 50 mM sodium acetate, and 2% (v/v) glycerol. The reaction was incubated at 55°C for 30 minutes and then chased with 10 mM Fe-BABE (Chapter 1) to a final concentration of 5 mM or 20 mM benzophenone maleimide (50% (v/v) DMF, Molecular Probes). The reactions were incubated at 37°C for 2 hours in the dark. Following derivatization, Fe-BABE modified pre-mRNAs were purified as described previously, benzophenone maleimide modified pre-mRNAs were extracted with phenol/chloroform/isoamyl alcohol, and ethanol precipitated with 2 μg glycogen. Derivatized pre-mRNAs were resuspended in distilled water and stored at -20°C.

Affinity cleavage. Affinity cleavage reactions were performed in 30 μL reactions containing 25% (v/v) HeLa nuclear extract or U2AF depleted extract that had been dialyzed into glycerol free buffer D (20 mM HEPES, pH 7.9, 0.1M KCl, 0.5 mM DTT) as described previously (Chapter 1). Reactions for E and E' complex (45 minutes at 30°C) contained 20 mM KCl, and 4 units ribonuclease inhibitor, but lacked ATP/CP and $MgCl_2$. Cleavage reactions were initiated by the addition of H_2O_2 and ascorbic acid as described previously (Chapter 1). Reactions were quenched with thiourea/glycerol (final concentrations 30 mM and 1% (v/v) respectively), treated with proteinase K, extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated.

UV crosslinking. Crosslinking reactions were performed in nuclear extract under E or E' complex conditions and irradiated with UV light on ice for 20 minutes. Complexes were immediately purified by native agarose gel electrophoresis as described above. Gels were visualized by autoradiography, and the isolated complexes excised from the gel. Crosslinked proteins where extracted from the gel slices with an electroelutor (200V for 3 hours), the RNA was digested in elution buffer (1×TBE) with 2 units of RNase A. Digestion reactions were incubated at 37°C for 60 minutes, and then reduced to approximately 20 μL on a speed-vac concentrator. Loading dye was added to each sample and subjected to 12% (29:1 acrylamide/bisacrylamide) SDS PAGE. Gels were visualized with coomassie blue, dried, and exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Primer extension. Primer extension reactions were performed as described previously (Chapter 1; Stern et al., 1988). For analysis of the 5' splice site the primer (5'-TCG AGA CGA GCT GAC ATC-3') complementary to the region immediately 5' of the branch site was used. For analysis of U1 snRNA the primer (5'-CAG GGG AAA GCG CGA ACG-3') complementary to the 3' end of U1 snRNA was used. For analysis of U2 snRNA the primer (5'-TGG TGC ACC GTT CCT GG-3') complementary to the 3' end of U2 snRNA was used. Primer extension reactions contained approximately 1-5 pmol RNA substrate and 1 μM DNA primer. Extension was carried out using Super Script RT (Gibco BRL). Following extension reaction were extracted with phenol/chloroform/isoamyl alcohol, ethanol precipitated, and then subjected to 8% (19:1 acrylamide/bisacrylamide) denaturing polyacrylamide gel electrophoresis. Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

3-4. References.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-412.

Abovich, N., Liao, X.C., and Rosbash, M. (1994). The yeast MUD2 proetin: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev.* **8**, 843-854.

Barabino, S.M.L., Blencowe, B.J., Ryder, U., Sproat, B.S., and Lamond, A.I. (1990). Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**, 293-302.

Bennett, M., Michaud, S., Kingston, J., and Reed, R. (1992). Protein components specifically associated with prespliceosomes and spliceosome complexes. *Genes Dev.* **6**, 1986-2000.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.

Blencowe, B.J., and Lamond, A.I. (1999). Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* **118**, 275-287.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the Spliceosome. In The RNA World 2nd Ed., R.F. Gesteland, T.R. Cech, J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525-560.

Burns, C.G., Ryoma, O., Krainer, A.R., and Gould, K.L. (1999). Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13789-13794.

Champion-Arnaud, P., Gozani, O., Palandjian, L., and Reed, R. (1995). Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. *Mol. Cell Biol.* **15**, 5750-5756.

Cote, J., Beaudoin, R.T., and Chabot, B. (1995). The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' splice site. *J. Biol. Chem.* **270**, 4031-4036.

Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**, 1504-1508.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86-90.

Kent, O.A., and MacMillan, A.M. (2002). Early Organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.* **9**, 576-581.

Kent, O.A., Reayi, A., Foong, L., Chilibeck, K.A., and MacMillan, A.M. (2003). Structuring of the 3' splice site by U2AF65. *J. Biol. Chem.* **278**, 50572-50577.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.

Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell Biol.* **8**, 3755-3760.

Li, Y., and Blencowe, B.J. (1999). Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *J. Biol. Chem.* **274**, 35074-35079.

MacMillan, A.M., Query, C.C., Allerson, C.A., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic Association of Proteins with the pre-mRNA Branch Region. *Genes Dev.* **8**, 3008-3020.

MacMillan, A.M., McCaw, P.S., Crispino, J.D., and Sharp, P.A. (1997). SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 133-136.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534-2546.

Michaud, S., and Reed, R. (1993). A functional association between the 5' and 3' splice sites is established in the earliest prespliceosome complex (E) in mammals. *Genes & Dev.* **7**, 1008-1020.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587-597.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**, 340-345.

Rutz, B., and Seraphin, B. (1999). Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* **5**, 819-831.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-358.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

Stern, S., Moazed, D., and Noller, H.F. (1988). Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481-489.

Wilson, K.S., and Noller, H.F. (1998). Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell* **92**, 131-139.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182-185.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-838.

Chapter 4.
Commitment complex organization in the absence of U1 snRNA-pre-mRNA base pairing.

Chapter 4. Commitment complex organization in the absence of U1 snRNA-pre-mRNA base pairing.

4-1. Introduction.

4-1.1. Association of U1 snRNP with the 5' splice site.

The earliest recognition of pre-mRNA substrates involves the ATP independent association of U1 snRNP and other non-snRNP associated factors with the pre-mRNA to form the commitment complex (yeast) or the early (E) complex in mammals (Burge et al, 1999; Krämer, 1996; Legrain et al., 1988; Michaud and Reed, 1991; Staley and Guthrie, 1998). In mammals, the non-snRNP associated factors include SF1 which recognizes the branch point region, U2AF containing large and small subunits which recognize the polypyrimidine tract and 3' splice site respectively, and members of the SR family of splicing factors (Berglund et al., 1997; Merendino et al., 1999; Staknis and Reed, 1994; Stark et al, 1998; Wu et al., 1999; Zahler et al., 1992; Zamore, et al., 1992; Zorio and Blumenthal, 1999). Previously, it has been demonstrated that U1 snRNP is a component of the commitment complex and formation of the commitment complex involved the stable association of U1 snRNP with the 5' splice site (Legrain et al., 1988; Serephin and Rosbash, 1989). We have identified a novel commitment complex (E') that forms in HeLa nuclear extract depleted of U2AF and have demonstrated that E' complex is a precursor to E complex in the mammalian pre-spliceosome assembly pathway (Chapter 3). The E' complex has the same sequence requirements for formation as the yeast CC1; this suggests that mammalian pre-spliceosome formation closely parallels the yeast system.

Mammalian U1 snRNP initiates pre-spliceosome assembly by binding the 5' splice site through base pairing interactions between the single stranded 5' terminal arm of U1 snRNA and a conserved stretch of six nucleotides at the 5' splice site of the pre-mRNA (Burge et al, 1999; Staley and Guthrie, 1998). Mammalian U1 snRNP consists of ten different proteins, and a 164 nucleotide uridine rich RNA (Stark et al., 2001). Seven of the proteins (the Sm proteins) are common to all snRNPs, whereas the other three (U1-70K, U1-A and U1-C) are specific to U1 snRNP. The Sm proteins are believed to form a ring structure around the Sm binding site of U1 snRNA. The proteins U1-70K and U1-A bind to the U1 snRNA stem loop I and II respectively (Stark et al., 2001; see Figure 4-2A). The smallest protein, U1 specific protein C (U1-C; 17.5 kDa), has not been shown to interact directly with the U1 snRNA, but probably associates with the U1 particle by protein-protein interactions. However, RNase footprinting experiments have demonstrated that U1-C protects the 5' arm of U1 snRNA (Tang et al., 1997).

The first step of pre-spliceosome assembly may be the association of U1 snRNP with the 5' splice site. It has been suggested that U1 protein•RNA interactions may precede the U1 snRNA•pre-mRNA base pairing interaction during early spliceosome assembly (Du and Rosbash, 2001). Studies with truncated U1 snRNA have demonstrated that U1 snRNP still associates with the 5' splice site and interaction of the 5' splice site with protein factors may be more important than base pairing for initial 5' splice site recognition. Recently, is has been shown by SELEX that U1-C can select a 5' splice site consensus sequence in which the first 4 nucleotides GUAU are identical to the first 4 nucleotides of the yeast 5' splice site consensus sequence (Du and Rosbash, 2002). The U1-C protein is conserved from yeast to mammals and therefore it has been proposed to play a fundamental role in pre-spliceosome assembly (Tang et al., 1997). For example, isolated U1 snRNPs depleted of U1-C do not associate with the pre-mRNA 5' splice site indicating a role of U1-C in the stable association of U1 snRNP (Heinrichs et al., 1990). Rosbash and coworkers suggest that a U1-C•5' splice site interaction precedes pre-mRNA•U1 snRNA base pairing and is the earliest step in the spliceosome assembly pathway (Du and Rosbash, 2002). EMSA experiments performed with a 5' splice site RNA pre-incubated with an antisense RNA centered on the 5' splice site demonstrated that U1-C did not interact with the duplex RNA.

In order to investigate 5' splice site recognition more closely, we carried out a series of experiments in U1 snRNA targeted RNase H digested nuclear extract to delineate the requirements for commitment of pre-mRNAs to pre-spliceosome assembly and splicing. We have determined that although truncated U1 snRNA extracts are unable to splice pre-mRNAs and unable to form commitment complexes, they are capable of sequestering pre-mRNAs in a commitment like fashion. However, in a doubly compromised extract, containing truncated U1 snRNA and lacking U2AF, this sequestering is not observed. This result suggests more than one mechanism for pre-mRNA commitment. For example, the interaction of U2AF with the polypyrimidine tract may act in concert with the U1 snRNP•pre-mRNA interaction to commit a pre-mRNA to splicing. We have probed pre-mRNAs in truncated U1 snRNA extract with the chemical nuclease Fe-BABE and detected a distinct cleavage pattern in the absence of U1 snRNA base pairing. This result suggests that the 5' splice site is initially held in proximity to the branch region by a protein•RNA interaction that precedes the canonical U1 snRNA•pre-mRNA interaction. We suggest that the initial protein•RNA interaction involves the U1-C•5' splice site interaction.

4-2. Results and Discussion.

4-2.1. U1 snRNP is a commitment complex component.

Previous studies have demonstrated that U1 snRNP is a major component of the commitment complex in both yeast and mammals (Legrain et al., 1988; Michaud and Reed, 1991; Serephin and Rosbash, 1989). In order to elucidate early assembly requirements of pre-spliceosome assembly, we specifically depleted HeLa nuclear extract of U1 snRNP using 2'-O-methyl-biotinylated oligonucleotides complimentary to the 5' arm of U1 snRNA (Barabino et al., 1990; Blencowe and Lamond, 1999) and analyzed the commitment ability of depleted extract (Figure 4-1). Wild type PIP85.B pre-mRNA were incubated in HeLa nuclear extract under E complex forming conditions and splicing analyzed when spliceosome assembly was challenged with an excess of competitor RNA (Figure 4-1A). Splicing of body-labeled pre-mRNA in WT extract is severely inhibited in the presence of a 500-fold molar excess of cold competitor RNA (Figure 4-1A; compare lanes 1 and 2). However, body-labeled pre-mRNA pre-incubated at 30°C to form the commitment complex still splices to normal levels when chased with extract saturated with a 500-fold molar excess of cold competitor pre-mRNA (Figure 4-1A; lanes 3 and 4).

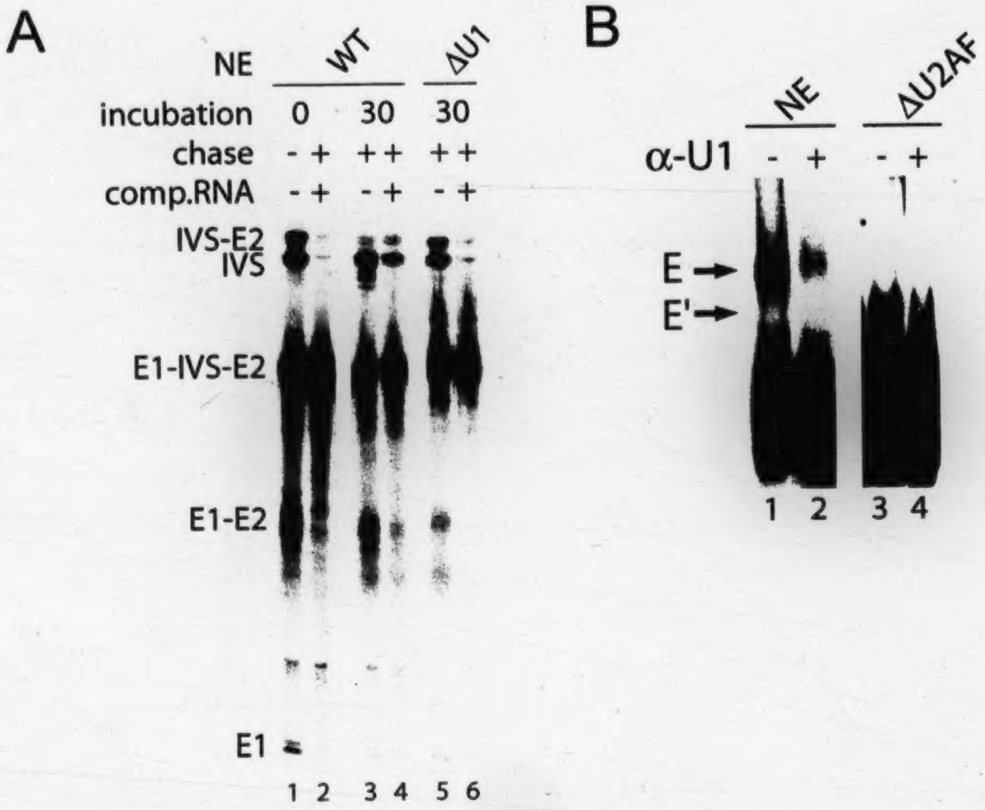


Figure 4-1. Analysis of U1 snRNP depleted and U1 snRNP blocked nuclear extracts. **(A)** U1 snRNP is required for pre-mRNA commitment to splicing. Pre-mRNAs were incubated at 0°C or 30°C in wild type (WT) or U1 snRNP depleted (ΔU1) nuclear extract under commitment complex forming conditions and chased with wild type extract in the presence (+) or absence (-) of 500-fold competitor RNA. Splicing intermediates and products indicated. **(B)** Native agarose gel shift analysis of complexes formed in nuclear extract (NE) and U2AF depleted (ΔU2AF) nuclear extract in the absence of U1 snRNA•pre-mRNA interaction. Commitment complex formation was analyzed in the absence (-) and presence (+) of a 2'-O-methyl biotinylated anti-U1 oligonucleotide.

We challenged commitment complex formation in extract specifically depleted of U1 snRNP. Normal levels of splicing products are detected when the body labeled pre-mRNA is pre-incubated under commitment complex forming conditions in U1 depleted extract and chased with wild type extract and ATP (Figure 4-1A; lane 5). This result indicates that the chase mix is sufficient to support splicing in the depleted extract. However, no splicing of the body labeled pre-mRNA is observed when the chase is saturated with cold competitor RNA (Figure 4-1A; lane 6). This confirmed previously reported data that U1 snRNP is required for commitment of pre-mRNAs to the splicing pathway.

We wanted to further investigate the role of the U1 snRNA•pre-mRNA interaction in commitment complex formation. In order to more fully examine an early role of U1 snRNA in E complex formation, we incubated wild type and U2AF depleted HeLa nuclear extract with a 2'-O-methyl-biotinylated oligonucleotide complementary to the 5' arm of U1 snRNA (see Figure 4-2A). This region has previously been determined to contact the 5' splice site. We challenged E complex and E' complex formation in the presence of the blocking oligo (Figure 4-1B) and used native agarose gel electrophoresis to visualize the complexes (Das and Reed, 1999). Neither the E complex nor the E' complex were detected by native agarose gel when the 5' arm of U1 snRNA was blocked and thus prevented from interacting with the pre-mRNA (Figure 4-1B; lanes 2 and 4). This result suggested that the U1 snRNA•pre-mRNA interaction is important for commitment complex formation. However, the inability to detect complexes on an agarose gel could be a result of an unstable complex and may not be a true reflection of the ability of U1 snRNP to associate with the pre-mRNA at this stage of pre-spliceosome assembly.

4-2.2. U1 snRNP association with the 5' splice site.

Although a U1 snRNA•pre-mRNA interaction is required for commitment complex identification by native agarose gel, it does not preclude the possibility that pre-mRNAs can be committed to splicing in the absence of this interaction. In order to further investigate the association of U1 snRNP with the pre-mRNA, we prepared nuclear extracts with a truncated U1 snRNA (Du and Rosbash, 2001). We decided to truncate the U1 snRNA by targeted RNase H digestion (Figure 4-2A). We synthesized a DNA oligonucleotide complementary to the first 12 nucleotides of U1 snRNA (U1-H, Figure 4-2A) and tested the ability of RNase H to digest off the end of U1 snRNA by incubating U1-H in both wild type and U2AF depleted nuclear extract. We analyzed the results by RT primer extension of the U1 snRNA with a primer complementary to the 3' end of U1 snRNA (Figure 4-2A). Incubation of 3 μM U1-H DNA oligonucleotide with RNase H in nuclear extract resulted in the complete digestion of U1 snRNA in both wild type and U2AF depleted extracts (Figure 4-2B; lanes 4 and 8). No cleavage of the U1 snRNA was observed when either extract was incubated with RNase H and no U1-H oligonucleotide (Figure 4-2B; lanes 2 and 5).

Although we had found that U1 snRNA is required for commitment complex formation, we wanted to determine if the 5' arm of U1 snRNA is accessible to RNase H digestion when it is involved in pre-mRNA interactions within commitment complexes. We assembled E and E' complexes on pre-mRNAs and then incubated the reaction with 3 μM U1-H oligonucleotide and RNase H. The RNAs were precipitated and analyzed by RT primer extension for U1 snRNA (Figure 4-2C). Following formation of either commitment complex, the 5' arm of U1 snRNA is not accessible to RNase H digestion (Figure 4-2C; lanes 4 and 7) as indicated by the presence of full length U1 snRNA analyzed by RT. However, if the U1-H oligonucleotide is co-incubated in the nuclear extract with the pre-mRNA substrate and RNase H, a competition is observed between U1 snRNA protection and U1 snRNA digestion. Based on gel quantification of the U1 snRNA primer extension products, approximately half of the U1 snRNA is protected against RNase H digestion (Figure 4-2C; lanes 3 and 6). This result confirms that association of U1 snRNP with the 5' splice site sequesters the 5' arm of U1 snRNA.

4-2.3. Splicing and complex analysis in truncated U1 snRNA extracts.

We tested the activity of truncated U1 snRNA extracts by conducting splicing reactions and analyzing complex formation (Figures 4-3A and 4-3B).

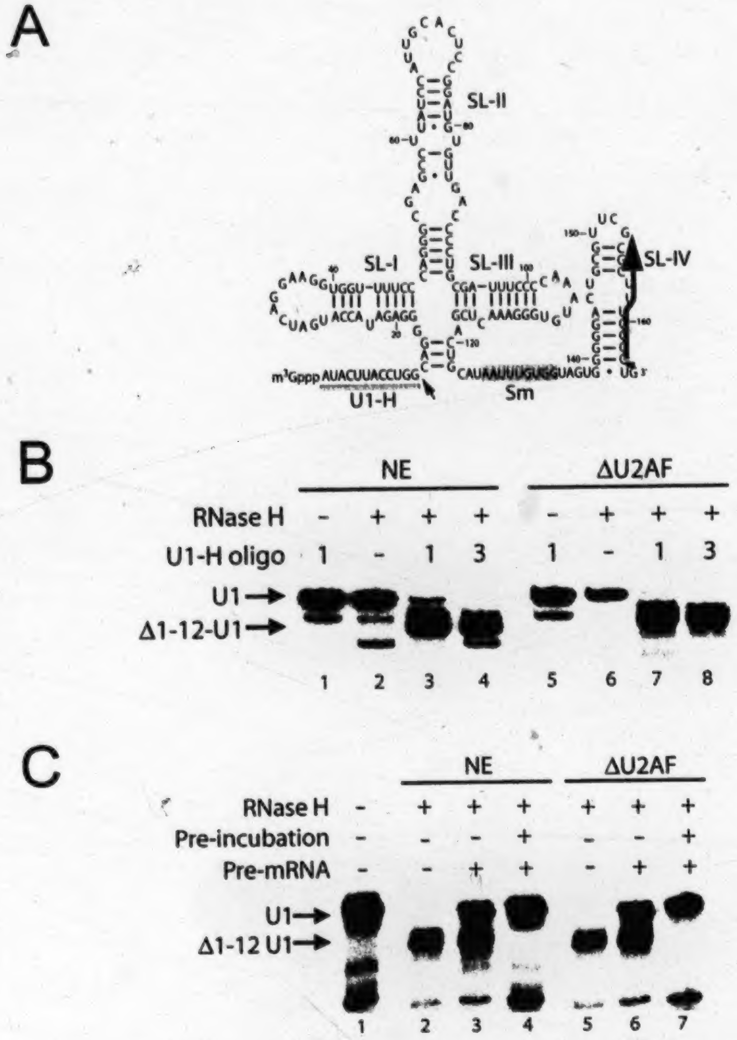


Figure 4-2. RNase H digestion of U1 snRNA. **(A)** Secondary structure of human U1 snRNA. U1 snRNA contains four stem loops (SL) a 5' arm structure with the 5' splice site binding sequence (yellow), and an Sm binding site (gray). For RNase H digestion a DNA oligonucleotide (U1-H) targeting the first 12 nucleotides (arrow) was used. For analysis of U1 snRNA by RT primer extension the DNA primer (green) complimentary to the 3' end was used. **(B)** RNase H digestion of U1 snRNA. Wild type (NE) or U2AF depleted (ΔU2AF) nuclear extract was incubated with U1-H (1 and 3 μM) and RNase H and digested at 30°C. U1 snRNA was analyzed by RT primer extension. **(C)** RNase H directed cleavage of U1 snRNA during commitment complex formation. Aliquots of wild type (NE) or U2AF depleted (ΔU2AF) nuclear extract were incubated without (-) or with (+) pre-mRNAs on ice (-) or after formation of the commitment complex (+) and probed with RNase H and U1-H DNA oligonucleotide. U1 snRNA was analyzed by RT primer extension.

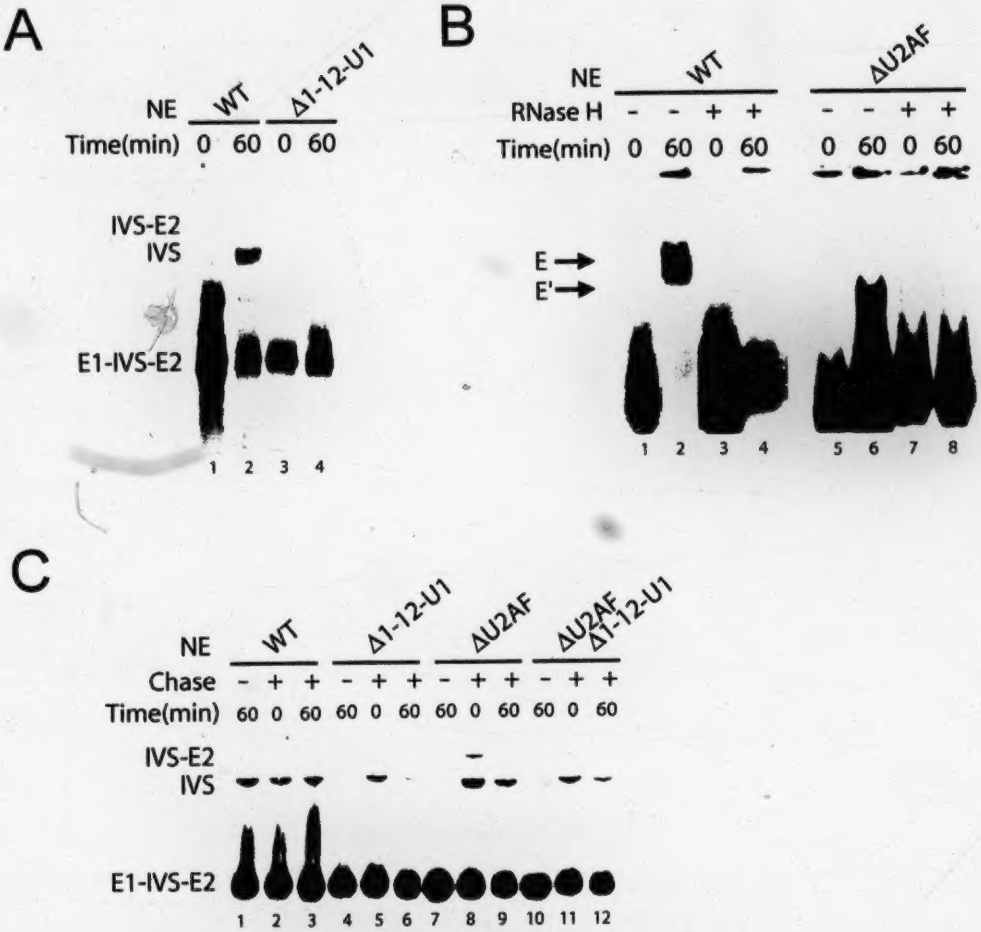


Figure 4-3. Sequestering pre-mRNA to pre-spliceosome complexes in the absence of U1 snRNA-pre-mRNA base pairing. **(A)** Truncated U1 snRNA nuclear extract is deficient for splicing. Pre-mRNAs were incubated in wild type (WT) or truncated U1 snRNA (Δ1-12-U1) nuclear extracts and analyzed for splicing. **(B)** Native agarose gel analysis of the formation of commitment complexes in RNase H digested nuclear extracts. Commitment complex formation was analyzed in wild type (WT) or U2AF depleted (ΔU2AF) nuclear extracts that had been digested with U1-H and RNase H. **(C)** Commitment of pre-mRNAs to splicing in the absence of U1 snRNA 5' end. Pre-mRNAs were incubated in the indicated extracts under commitment complex forming conditions (-$MgCl_2$/ATP) at 30°C for 60 minutes without a chase (lanes 1, 4, 7, and 10) on ice with a chase (lanes 2, 5, 8, and 11), or at 30°C for 60 minutes with a chase (lanes 3, 6, 9, and 12). Chase mix contained wild type extract, $MgCl_2$ and ATP. Following the indicated pre-incubation all reactions were further incubated at 30°C for 60 minutes and analyzed for splicing.

U1 targeted RNase H digested nuclear extract is significantly deficient in splicing activity (Figure 4-3A). Comparison of a 60 minute time point in wild type extract to a 60 minute time point in truncated U1 snRNA extract reveals a large decrease in splicing products (Figure 4-3A; lanes 2 and 4). The inability of truncated U1 snRNA extract to splice pre-mRNA is probably due to the inability of U1 snRNA to base pair to the 5' splice site. Commitment complex formation was not detected in either wild type or U2AF depleted extracts that had been digested with RNase H (Figure 4-3B). This confirms our previous data that commitment complex formation requires the presence of the 5' arm of U1 snRNA in order to form a stable complex able to be detected by native agarose gel electrophoresis.

Since U1 snRNP can associate with the pre-mRNA in the absence of base pairing via the interaction of U1-C with the pre-mRNA (Du and Rosbash, 2002), we wanted to investigate the ability of this association to commit the pre-mRNA to splicing. In order to indirectly investigate the commitment ability of U1 snRNP in truncated U1 extracts, pre-mRNAs were pre-incubated under E complex forming conditions before being chased with an aliquot of wild type extract and ATP (Figure 4-3C). Wild type extract was used as a control for the amount of splicing observed under these conditions (Figure 4-3C; lane 1). As expected no splicing was observed in truncated U1 snRNA extract, U2AF depleted extract, or the truncated U1-U2AF depleted extract (Figure 4-3C; lanes 4, 7, and 10). Similarly, when pre-mRNAs were incubated under E complex forming conditions in truncated U1 snRNA extract and chased with wild type extract a significant decrease in splicing was observed (Figure 4-3C; lane 6). This result suggests that the pre-mRNA is sequestered by pre-spliceosome factors in the absence of U1 snRNP-5' splice site base pairing preventing exchange of truncated U1 snRNP for wild type U1 snRNP. The interaction of U1 snRNP containing a truncated U1 snRNA with the 5' splice site may represent a "committed" pre-mRNA•splicing complex such that exchange with wild type U1 snRNP from the chase extract and thus splicing are not observed. The absence of splicing products suggests that the pre-mRNA is trapped in a complex containing the truncated U1 snRNA and is unable to proceed in spliceosome assembly. This interaction is probably between the 5' splice site and U1-C and suggests that the interaction of U1 snRNP with the 5' splice site via U1-C precedes the base pairing of U1 snRNA with the 5' splice site.

Pre-mRNAs incubated under E' complex forming conditions and chased with wild type extract splice to normal levels (Figure 4-3C; lane 9) suggesting that the E' complex is a biologically active complex and addition of U2AF to this complex allows spliceosome assembly to proceed. Unexpectedly, a low level of splicing was observed when the pre-mRNA was pre-incubated in truncated U1 snRNA-U2AF depleted extract and chased with wild type extract (Figure 4-3C; lane 12). This result suggests that U2AF may also play a more significant role in commitment of pre-mRNA to splicing. The presence of a low level of splicing products in the doubly compromised extract suggests that U1-C may be sampling the 5' splice site and require either U1 snRNA to become stably associated with the 5' splice site or U2AF to bind the polypyrimidine tract in order to commit the pre-mRNA to that complex. In the absence of both of these interactions, as in the doubly compromised extract, U1-C may be interacting with the 5' splice site, but may be in equilibrium between bound and unbound. Upon addition of the wild type chase extract, the wild type U1 snRNP is able to compete for 5' splice site usage with the truncated U1 snRNP and thus some splicing was observed.

4-2.4. Two conformations at the 5' splice site.

Previously, we probed the structure of RNA bound in both the E complex and the E' complex using pre-mRNA site-specifically derivatized with a directed hydroxyl radical probe, Fe-BABE, tethered to the pre-mRNA (Chapter 1; Chapter 3; Kent and MacMillan, 2002; Kent and MacMillan, submitted). Using a wild type RNA construct that had been derivatized with Fe-BABE at the branch region (WT148; Figure 4-4A) we probed complexes formed in wild type and U2AF depleted extracts in the presence of an antisense -U1 snRNA blocking oligonucleotide (Figure 4-4B) and U1 snRNA truncated NE (Figure 4-4C). Reactions were incubated under E' or E complex forming conditions (lacking ATP and $MgCl_2$) followed by Fe-BABE mediated cleavage and analysis of cleavage products by RT primer extension.

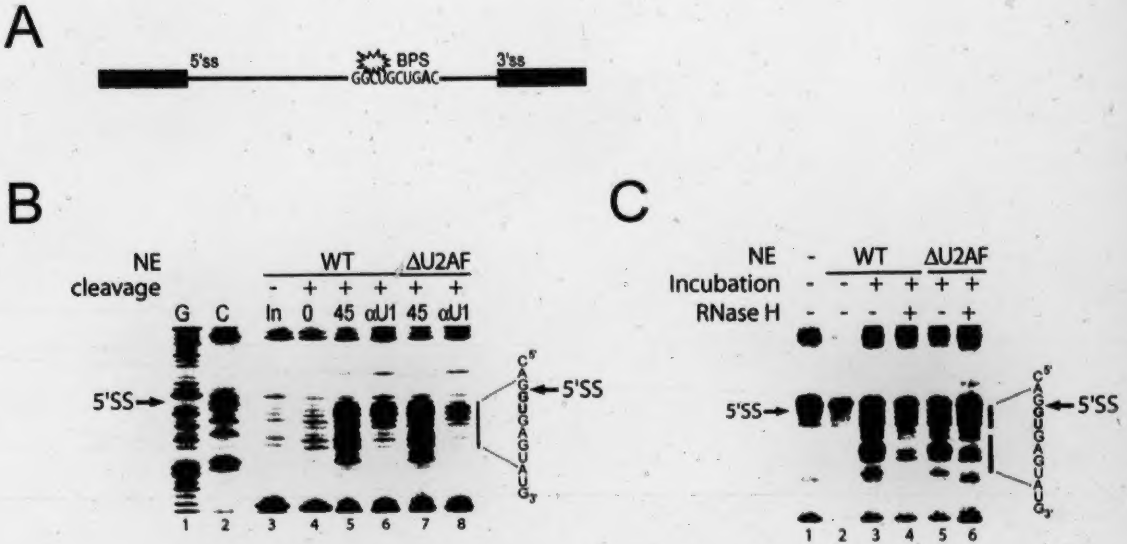


Figure 4-4. ***Directed hydroxyl radical cleavage in 1-12-U1 snRNA pre-spliceosome complexes.*** **(A)** Pre-mRNA substrate derivatized with Fe-BABE (star) at the branch region (WT148). **(B)** Analysis of cleavage reactions performed in nuclear extract (WT) and U2AF depleted (ΔU2AF) nuclear extract probed with Fe-BABE WT148 pre-mRNAs. Reactions were probed after incubation at 30°C for 0 or 45 minutes, with and without an anti-U1 snRNA blocking oligonucleotide (αU1). **(C)** Two cleavage patterns at the 5' splice site. Analysis of cleavage reactions performed in nuclear extract (WT) and U2AF depleted (ΔU2AF) nuclear extract digested with RNase H directed at the 5' end of U1 snRNA and probed with Fe-BABE WT148 pre-mRNAs. Indicated reactions were incubated under commitment complex forming conditions (+). RT reactions were compared to input RNA (In) for location of RT stops. G and C sequencing lanes shown. Sequence around the 5' splice site indicated. Regions of significant cleavage represented by vertical bars.

As reported previously (Chapter 1 and Chapter 3), when the pre-mRNA was probed in wild type and U2AF depleted extracts, cleavages are observed spanning the entire 5' splice site region (Figure 4-4B; lanes 5 and 7). Specifically, the cleavage was observed over 10 nucleotides at the 5' splice site. In addition the cleavage pattern resembles two overlapping cleavage regions suggesting that the 5' splice site under these conditions is in a helical conformation with the branch region on one side of the helix. Specifically, this result suggests that the 5' splice site is based paired to U1 snRNA in these complexes. Interestingly, if the pre-mRNA is probed in either extract pre-blocked with an anti-U1 oligonucleotide, a diminished cleavage pattern is observed (Figure 4-4B; lanes 6 and 8). The new pattern spans only 5 nucleotides and contains the 5' most region of the previously mentioned cleavage pattern. This suggests that in the absence of U1 snRNA base pairing the 5' splice site is held in a different conformation with respect to the branch point.

In order to further characterize this altered 5' splice site•branch region interaction, we probed the pre-mRNA in either wild type or U2AF depleted extract that had been digested with RNase H (Figure 4-4C). In complexes formed and probed in the truncated U1 snRNA extracts, only the minimal cleavage pattern is observed (Figure 4-4C; lanes 4 and 6). The minimal cleavage pattern may be from the 5' splice site region held in a single stranded conformation. This conformation is probably due to a protein-RNA interaction: specifically, the altered conformation may reflect the U1-C•5' splice site interaction.

4-2.5. U1-C interaction precedes U1 snRNA base pairing.

We have observed that commitment of pre-mRNAs to the splicing pathway occurs in the absence of U1 snRNA•pre-mRNA base pairing. U1 snRNP is likely the first splicing factor to specifically interact with the pre-mRNA substrate and is able to commit the pre-mRNA to splicing with minimal participation of other splicing factors (Legrain et al., 1988). Our data suggest that pre-spliceosome formation begins with an initial recognition of the 5' splice site probably involving the U1-C protein, followed by the canonical U1 snRNA base pairing interaction (Figure 4-5A).

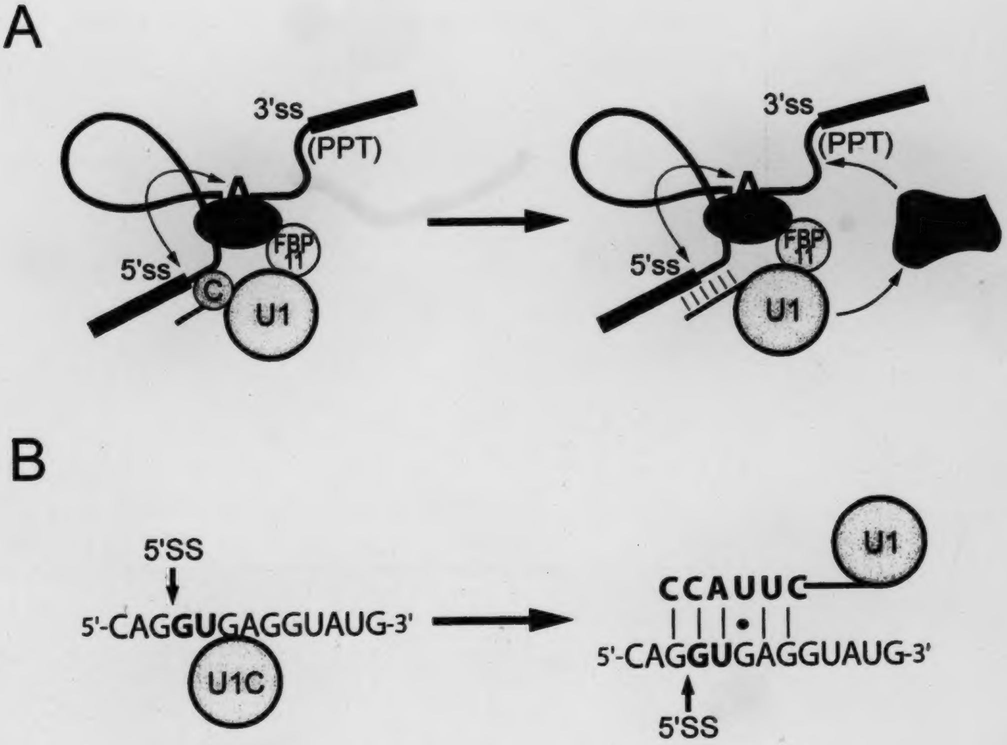


Figure 4-5. Two conformations at the 5' splice site. **(A)** Pre-spliceosome assembly proceeds through early recognition of the 5' splice site by the U1 snRNP specific protein U1-C followed by U1 snRNA-pre-mRNA base pairing and recruitment of U2AF to the polypyrimidine tract. Pre-mRNAs are committed to pre-spliceosome assembly by the association of U1 snRNP with the 5' splice site via the U1-C protein. SF1 interacts with the branch point region and U1 snRNP via the protein FBP11, resulting in the close proximity of the BPS and 5' splice site (green arrows). **(B)** Before U1 snRNA base pairing, the 5' splice site is held in a single stranded RNA•protein structure with U1-C (left). U1-C may potentiate the U1 snRNA-5' splice site base pairing structure observed with the formation of E' complex (right). U1 snRNP and/or SF1 recruit U2AF to the polypyrimidine tract resulting in E complex formation.

Previously, it has been suggested that U1-C may potentiate U1 snRNA base pairing by holding the U1 snRNA 5' arm in a conformation poised for base pairing with the pre-mRNA substrate (Tang et al., 1997). It has been suggested that base pairing between U1 snRNA and the 5' splice site is probably required for driving subsequent steps in spliceosome assembly (Du and Rosbash, 2001). For example, splicing is not observed in U1 snRNA blocked or U1 depleted extracts. There may be a conformational change that promotes the transition from E complex to active spliceosome. Therefore, the U1 snRNA truncated complex is not a functional complex but represents a freeze frame in pre-spliceosome formation.

Previously, we have observed a close proximation of the branch region•5' splice site in E and E' complex using pre-mRNA probes modified with Fe-BABE (Kent and MacMillan, 2002, Kent and MacMillan, submitted). In the present study, we have determined that pre-mRNAs incubated under E complex forming conditions in the absence of U1 snRNA•pre-mRNA base pairing are structured to position the branch region near the 5' splice site. However, in the absence of U1 snRNA base pairing a diminished cleavage pattern was observed. The interaction of U1 snRNP with the pre-mRNA in this complex probably reflects the 5' splice site bound by U1-C (Figures 4-5A and 4-5B; left). Upon U1-C binding, the U1 snRNA is base-paired to the 5' splice site to form the E' complex (Figure 4-5A; right) which results in the extended Fe-BABE cleavage pattern due to the pre-mRNA 5' splice site being in a helical conformation with U1 snRNA (Figure 4-5B; right). The close proximation of the 5' splice site and branch region to one another in these complexes probably reflects bridging interactions between U1 snRNP and the branch binding protein SF1 (Figure 4-5A). The interactions between SF1, the mammalian Prp40-like protein FBP11, and U1 snRNP have been proposed to bridge the 5' splice site and the branch region in E complex (Abovich and Rosbash, 1997; Reed, 2000) and may be present in complexes that precede U1 snRNP•pre-mRNA base pairing.

4-3. *Materials and Methods.*

4-3.1. *Synthesis of pre-mRNA substrates.*

Plasmid preparation. The PIP85.B plasmid containing the PIP85.B pre-mRNA and an ampicillin resistance cassette was transformed into competent *E.coli* BL21 cells and purified as described previously (Chapter 1).

PCR of transcription templates. Templates for transcription were made by PCR using the PIP85.B plasmid as the template DNA. The primers M13F and CQ27 were used to create a PCR product containing a T7 RNA polymerase promoter representing a transcription template for the full length PIP85.B pre-mRNA. Plasmid and primers were incubated in reactions that contained 10 mM Tris-HCl, pH 9, 1.5 mM $MgCl_2$, 50 mM KCl, 1 mM of each dNTP and 2 units TAQ DNA polymerase as described previously (Chapter 1).

Transcription. RNAs were transcribed from PCR products derived from the PIP85.B plasmid described above. Standard transcription reactions contained 40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM each ATP, CTP, and GTPs, 15 μM UTP, 100 μCi α-[^{32}P]-UTP (3000 Ci/mmol, 800 μCi/uL, NEN), 1 mM DTT, 100 ng PCR derived template, and 2 units RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, quenched with loading dye and loaded onto 8% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. Standard transcription conditions were used to create unlabeled pre-mRNAs. Full length pre-mRNA (13/27) was transcribed from the PCR product in reactions that contained 40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM each NTP, 10 mM DTT, 100-400 ng PCR derived template, and 10 units RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, centrifuged to remove insoluble inorganic phosphate accumulated during the reaction, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. Body-labeled RNAs were visualized by autoradiography using Kodac X-OMAT film, unlabeled RNAs were visualized by UV shadowing, excised from the gel and further purified as previously described (Chapters 1 and 3).

Site specifically modified pre-mRNAs. RNAs that contained the N4-functionalized cytosine were prepared as described (Appendix *II*). An oligonucleotide representing the PIP85.B pre-mRNA branch point sequence (5'-GG(ClϕU)UGCUGAC-3') was incorporated into full length pre-mRNAs using a three piece ligation procedure with 5' and 3' pieces of RNA and a DNA bridging oligonucleotide as described previously (Chapter 1).

4-3.2. Analysis of pre-mRNA splicing and complex formation.

Preparation of RNase H digested nuclear extract. Wild type and U2AF depleted nuclear extracts were digested with RNase H in the presence of a DNA oligonucleotide directed at the 5' end of U1 snRNA. Digestion reactions (10 μL) contained 25% (v/v) nuclear extract, 2 units of RNase H (Roche), and 3 μM U1-H DNA template (5'-CCA GGT AAG TAT-3'). Reactions were incubated at 30°C for 30 minutes. Following digestion, reactions were digested with proteinase K, phenol/chloroform/isoamyl alcohol extracted, and ethanol precipitated. Precipitated RNAs were subjected to RT primer extension (see below), ethanol precipitated, and run on 8% (19:1 acrylamide /bisacrylamide) denaturing PAGE. Large scale preparations of digested extract were prepared in an analogous fashion, however, following digestion aliquots of the digested nuclear extract were flash frozen and stored at -80°C.

Pre-spliceosome complex formation. For analysis of E complex and E' complex formation, pre-mRNAs ($50\text{-}100 \times 10^3$ c.p.m.) were incubated in reactions containing 25% (v/v) HeLa nuclear extract or 25% (v/v) U2AF depleted HeLa nuclear extract respectively, 60 mM KCl, and 4 units RNase inhibitor. Reactions were incubated at 30°C for 45 to 120 minutes, stopped by the addition of 97% (v/v) glycerol loading dye, and were fractionated on 1.2% (w/v) agarose, 0.5×TBE gels (Chapter 3). Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular

Dynamics Storm 840 Phosphorimager. Specific interactions of U1 snRNP and U2 snRNP with the pre-mRNA were blocked with anti-U1 and anti-U2 biotinylated-2'-O-methyl anti-sense oligonucleotides (Dharmacon) as described elsewhere (Chapter 1).

RNase H protection assay. Pre-mRNAs were incubated in wild type or U2AF depleted nuclear extracts under E or E' complex forming conditions and incubated at 30°C for 60 minutes. Following incubation, RNase H (2 units) and 3 μM U1-H DNA oligonucleotide were added and the reaction allowed to incubate at 30°C for an additional 30 minutes. Following digestion, reactions were purified and RT primer extended for U1 snRNA as described above.

Splicing assays. Pre-mRNAs ($50\text{-}100 \times 10^3$ c.p.m.) were incubated in 20 μL reactions containing 40% (v/v) HeLa nuclear extract, 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and 4 units ribonuclease inhibitor. Commitment assays were performed by incubating pre-mRNA ($50\text{-}100 \times 10^3$ c.p.m.) under E or E' complex conditions as described above. Aliquots were then added to a splicing mix containing ATP/CP and $MgCl_2$ with or without competitor RNA (500 pmol). Following incubation at 30°C, splicing reactions were digested with 40 μg proteinase K in the presence of 20 μg tRNA at 55°C for 20 minutes as described previously (Chapter 1). Reactions were extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated, and then subjected to 15% (19:1 acrylamide/bisacrylamide) 8 M urea denaturing polyacrylamide gel electrophoresis. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Commitment assays. Pre-mRNAs ($50\text{-}100 \times 10^3$ c.p.m.) were incubated in 20 μL reactions containing 25% (v/v) HeLa nuclear extract under E or E' complex conditions in wild type and U1 snRNA digested nuclear extracts as described above. After incubation at 30°C for 60 minutes reactions were chased with splicing mix containing 25% (v/v) nuclear extract, 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 5 mM creatine phosphate, and 4 units ribonuclease inhibitor. Commitment assays were performed by incubating pre-mRNA (50-100 × 10^3 c.p.m.) under E or E' complex conditions as described above. Aliquots were then added to a splicing mix containing ATP/CP and $MgCl_2$ with or without competitor RNA. Reactions were digested with proteinase K, phenol/chloroform extracted, and ethanol precipitated as previously described. Reactions were subjected to 15% (19:1 acrylamide/bisacrylamide) 8 M urea denaturing polyacrylamide gel electrophoresis. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Derivatization of N4-functionalized cytosine. Pre-mRNA ligation products containing the N4-functionalized cytosine at the branch region were derivatized with Fe-BABE under reducing conditions and purified as described previously (Chapter 1). Derivatized pre-mRNAs were resuspended in distilled water and stored at -20°C.

4-3.3. Analysis of Fe-BABE cleavage reactions.

Affinity Cleavage. Affinity cleavage reactions were performed in 30 µL reactions containing 25% (v/v) HeLa nuclear extract or U2AF depleted extract that had been dialyzed into glycerol free buffer D (20 mM HEPES, pH 7.9, 0.1M KCl, 0.5 mM DTT) as described previously (Chapter 1). Reactions for E and E' complex (45 minutes at 30°C) contained 20 mM KCl, and 4 units ribonuclease inhibitor, but lacked ATP/CP and $MgCl_2$. Cleavage reactions were initiated by the addition of H_2O_2 and ascorbic acid as described previously (Chapter 1). Reactions were quenched with thiourea/glycerol (final concentrations 30 mM and 1% (v/v) respectively), treated with proteinase K, extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated.

Primer Extension. Primer extension reactions were performed as described previously (Chapter 1, Stern et al., 1988). For analysis of the 5' splice site the primer (5'-TCG AGA CGA GCT GAC ATC-3') complementary to the region immediately 5' of the branch site was used. The primer (5'-CAG GGG AAA GCG CGA ACG-3') was used for analysis of U1 snRNA in RNase H treated assays. Primer extension reactions contained approximately 1-5 pmol RNA substrate and 1 μM DNA primer. Extension was carried out using Super Script RT (Gibco BRL). Following extension, reactions were extracted with phenol/chloroform/isoamyl alcohol and ethanol precipitated, and then subjected to 8% (19:1 acrylamide/bisacrylamide) 8 M urea denaturing PAGE. Dried gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

4-4. References.

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-412.

Barabino, S.M.L., Blencowe, B.J., Ryder, U., Sproat, B.S., and Lamond, A.I. (1990). Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**, 293-302.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.

Blencowe, B.J., and Lamond, A.I. (1999). Purification and depletion of RNP particles by antisense affinity chromatography. *Methods Mol. Biol.* **118**, 275-287.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the Spliceosome. In The RNA World 2nd Ed., R.F. Gesteland, T.R. Cech, J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525-560.

Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**, 1504-1508.

Du, H., and Rosbash, M. (2001). Yeast U1 snRNP-pre-mRNA complex formation without U1 snRNA-pre-mRNA base pairing. *RNA* **7**, 133-142.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86-90.

Heinrichs, V., Bach, M., Winkelmann, G., and Luhrmann, R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, **247**, 69-72.

Kent, O.A., and MacMillan, A.M. (2002). Early Organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.* **9**, 576-581.

Kent, O.A., and MacMillan, A.M. (2003). Characterization of a U2AF independent commitment complex (E') in the mammalian spliceosome assembly pathway. *Submitted.*

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Ann. Rev. Biochem.* **65**, 367-409.

Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol. Cell Biol.* **8**, 3755-3760.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534-2546.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**, 340-345.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-358.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell Biol.* **14**, 7670-7682.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.

Stark, J.M., Bazett-Jones, D.P., Herfort, M., and Roth, M.B. (1998). SR proteins are sufficient for exon bridging across an intron. *Proc. Natl. Acad. Sci. USA* **95**, 2163-2168.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.

Stern, S., Moazed, D., and Noller, H.F. (1988). Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481-489.

Tang, J., Abovich, N., Fleming, M., Serephin, B., and Rosbash, M. (1997). Identification and characterization of the yeast homologue of U1 snRNP-specific protein C. *EMBO J.* **16**, 4082-4091.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837-847.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* **402**, 835-838.

Chapter 5.

Future directions: transcription and RNA processing.

Chapter 5. Future directions: transcription and RNA processing.

5-1. Future Directions.

The techniques outlined in this thesis could be extended to gain a more complete understanding of spliceosome assembly and in addition study the cooperative interactions between pre-mRNA splicing, transcription and other RNA processing events such as 5' cap formation and A to I editing.

5-1.1. Applications to spliceosome assembly.

The thesis has outlined techniques that have been used to characterize the structural organization of pre-mRNA substrates within the mammalian splicing commitment (E) complex. The Fe-BABE approach (Chapter 1) could be further developed to characterize RNA-RNA rearrangements during the formation of the active spliceosome. Although the interactions of the snRNAs during spliceosome assembly have been well characterized (Nilsen, 1998), the order of the snRNA rearrangements with respect to each other is yet to be determined. Possible experiments with pre-mRNA tethered Fe-BABE may help to elucidate the temporal arrangement of both the snRNAs and the pre-mRNA during the formation of the spliceosome active site. In addition, this technique could be used to determine the order of helix dissociations and helix formations that occur in the rearrangement of the spliceosomal snRNAs prior to both the first and second steps of splicing.

Although the presence of ATP is required for the stable association of U2 snRNA with the branch region in complex A, it is unclear how U2 snRNP is recruited to E complex and subsequently associated with the branch region. A protein-protein interaction between SAP155, a U2 snRNP component, and U2AF65 has been detected by immunoprecipitation experiments suggesting that U2AF may recruit U2 snRNP via an interaction with SAP155 (Das et al., 2000). In addition, the RS domain of U2AF65 has been shown to be essential to spliceosome assembly and associates with the pre-mRNA branch region (Valcarcel et al., 1996). N-terminal U2AF65 deletion mutants (Chapter 2) modified with an N-terminal photocrosslinker could be used in experiments performed during pre-spliceosome formation. These experiments may reveal the interactions of the U2AF65 RS domain with factors required for U2 snRNP addition to E complex and ultimately branch region association.

Crosslinking experiments using site specifically modified pre-mRNAs have identified the presence of a U2 snRNP protein, p14, which associates with the pre-mRNA branch region at all stages of spliceosome assembly (MacMillan et al., 1994). Currently, the biological role of p14 is unknown. We have performed preliminary pull-down experiments that demonstrate an interaction of p14 with the E complex component U2AF65 (Schellenberg, Kent, and MacMillan, unpublished). Specifically, p14 has been demonstrated to interact with both the region between the linker and RRM1 and with a region of RRM3 in U2AF65. These interactions suggest a possible mechanism by which U2AF via several protein-protein interactions may recruit U2 snRNP to the branch region to form complex A. Experiments based on procedures outlined in this thesis may help to elucidate the mechanism by which U2 snRNP is recruited to E complex and how U2 snRNA is ultimately based paired to the branch region.

5-1.2. Transcription and pre-mRNA splicing are coordinated events.

The expression of eukaryotic genes begins with nuclear DNA-dependent RNA polymerase (RNAP) that binds to a promoter and directs synthesis of a nascent RNA (Lee and Young, 2000). Of the three RNA polymerases, RNAP II transcribes the broadest variety of sequences and is solely responsible for the synthesis of pre-mRNAs (Howe, 2002). RNA polymerase II contains an unusual domain at the carboxyl terminus termed the CTD that has been extensively characterized (Allison et al., 1985; Corden, 1990; Meredith et al., 1996). Although the CTD has been implicated in all stages of the transcription cycle (Howe, 2002), of particular relevance to the work described in this thesis is the role of the CTD in coordinating RNA processing. Specifically, the CTD has been implicated in 5' cap formation, pre-mRNA splicing, and poly-adenylation (Cramer et al., 2001).

Electron microscopy analyses have demonstrated that pre-mRNA splicing is co-transcriptional (Beyer and Osheim, 1988) and more recent reports have demonstrated that splicing factors, such as those of the SR protein family, associate with specific regions of the CTD domain to direct spliceosome assembly to sites of active transcription (Proudfoot, 2000). Furthermore, the CTD domain has been shown to interact with the capping machinery required for 5' end formation. Capping occurs early during RNA transcription and the acquisition of the 5' cap structure is believed to act as a switch from transcriptional initiation to elongation (McCracken et al., 1997). Therefore, the CTD is thought to act as a platform that facilitates the assembly of complexes required for processing pre-mRNAs (Rosonina and Blencowe, 2004).

Future experiments may help to define the role of the CTD in recruiting constitutive splicing factors to 5' splice sites and provide an understanding of how and if the multifunctional CTD participates in commitment of pre-mRNA to splicing. Specifically, the recruitment of U1 snRNP to the 5' splice site has been determined to represent the earliest step in sequestering of pre-mRNA into pre-spliceosomal complexes (Chapter 3 and Chapter 4) but how U1 snRNP is recruited to sites of active transcription is not clear. Since the capping enzymes are also associated with the CTD (Cho et al., 1997) and cap binding complex interacts with components of U1 snRNP (Lewis et al., 1996; Zhang and Rosbash, 1999) this may suggest a possible mechanism for recruitment of U1 snRNP to nascent 5' splice sites. In addition, U1 snRNP can be recruited to the pre-mRNA 5' splice site via interactions with SR proteins such as ASF/SF2, which may also interact with the CTD of RNAP II (Eperon et al., 2000; Proudfoot, 2000).

5-1.3. Pre-mRNA splicing and A to I RNA editing.

RNA editing by members of the ADAR (adenosine deaminases that act on RNA) family involves hydrolytic deamination of adenosine to inosine within the context of double-stranded RNA (Bass, 1992). A specific example of A to I editing occurs in the pre-mRNA of the neural-specific AMPA class of glutamate-gated ion channels at the Q/R and R/G editing sites, which ultimately affects the properties of the resulting channels (Sommer et al., 1991). The Q/R and R/G editing sites are both located within exons close to 5' splice sites. The exonic sequence surrounding the R/G site for example requires downstream intron sequences in order to form the double stranded RNA structure required for editing (Higuchi et al., 1993). This structure sequesters the 5' splice site of the intron in a double-stranded structure that is inaccessible to U1 snRNP and spliceosome assembly. Therefore, editing of the R/G site occurs before splicing. However, a link between editing by the ADARs and the pre-mRNA splicing machinery has not been established. Previous work has demonstrated that ADAR has an inhibitory effect on splicing of model substrates derived from the R/G pre-mRNA which can be partially relieved by RNA helicase A (Bratt and Öhman, 2003). Presumably, the helicase unwinds the double-stranded structure exposing the 5' splice site to splicing machinery; however, *in vivo* the same RNA construct was found to edit and splice without interference suggesting that the two processing events are coordinated (Bratt and Öhman, 2003).

Future experiments may help to uncover a link between A to I editing by the ADARs and pre-mRNA splicing. For example, analysis of pre-spliceosome commitment complexes (Chapter 3) on substrates derived from the R/G pre-mRNA may demonstrate a requirement for ADARs in formation of the E complex. Furthermore, analysis of U1 snRNP recruitment to the 5' splice site region of R/G pre-mRNAs may uncover an interaction of U1 snRNP components with ADAR. A potential interaction may be direct or could possibly involve members of the SR family of splicing factors (Gravely, 2000).

5-2. References.

Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599-610.

Bass, B. (1992). In *The RNA World* (eds Gesteland, R.F., Cech, T.R., and Atkins, J.F.) 383-418 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754-765.

Bratt, E., and Öhman, M. (2003). Coordination of editing and splicing of glutamate receptor pre-mRNA. *RNA* **9**, 309-318.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11**, 3319-3326.

Corden, J.L. (1990). Tails of RNA polymerase II. *Trends Biochem Sci.* **15**, 383-387.

Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogues, G., and Kornblihtt, A.R. (2001). Coordination between transcription and pre-mRNA processing. *FEBS Lett.* **498**, 179-182.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol. Cell.* **5**, 779-787.

Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Caceres, J.F., Hayward, D.G., and Krainer, A.R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell Biol.* **22**, 8303-8318.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA* **9**, 1197-1211.

Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., and Seeburg, P.H. (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361-1370.

Howe, K.J. (2002). RNA polymerase II conducts a symphony of pre-mRNA precessing activities. *Biochim. Biophys. Acta* **1577**, 308-324.

Lee, T.I, and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77-137.

Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I.W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* **10**, 1683-1698.

MacMillan, A.M., Query, C.C., Allerson, C.A., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic Association of Proteins with the pre-mRNA Branch Region. *Genes Dev.* **8**, 3008-3020.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361.

Meredith, G.D., Chang, W.H., Li, Y., Bushnell, D.A., Darst, S.A., and Kornberg, R.D. (1996). The C-terminal domain revealed in the structure of RNA polymerase II. *J. Mol. Biol.* **258**, 413-419.

Nilsen, T.W. (1998). RNA-RNA Interactions in Nuclear Pre-mRNA Splicing. In *RNA Structure and Function*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), 279-307.

Proudfoot, N. (2000). Connecting transcription to messenger RNA processing. *Trends Biochem Sci.* **25**, 290-293.

Rosonina, E. and Blencowe, B.J. (2004). Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *RNA* **10**, 581-589.

Sommer, B., Kohler, M., Sprengel, R., and Seeburg, P.H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11-19.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* **273**, 1706-1709.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.

Appendix *I*.

Kinetic analysis of the M1 RNA folding pathway.

Appendix *I*. Kinetic analysis of the M1 RNA folding pathway (1).

I-1. Introduction.

I-1.1. RNase P M1 RNA structure and folding.

RNase P is the essential and ubiquitous endonuclease required for the maturation of the 5' end of precursor-tRNAs (ptRNAs). RNase P containing both RNA and protein component cleaves the 5' leader sequence from ptRNAs to generate the mature tRNA (Figure *I*-1A). Although the activity consists of both protein and RNA components, in some cases the RNA alone is sufficient for site-specific pre-tRNA cleavage (Guerrier-Takada et al., 1983; Altman et al., 1999). For example, the RNase P activity from *Escherichia coli* is a complex of the M1 RNA and C5 protein, but in the presence of high concentrations of either divalent magnesium or salt, M1 RNA alone is able to efficiently cleave ptRNA (Guerrier-Takada et al., 1983). The RNA components of *E. coli* RNase P, and the related RNA from *Bacillus subtilis*, are large molecules (377 and 408 nucleotides respectively) that require divalent magnesium to fold into their active structures and are thus good model systems to study RNA folding.

The secondary structure of the M1 RNA, predicted on the basis of phylogenetic comparisons (James et al., 1988; Westhof and Altman, 1994; Massire et al., 1998), contains a series of paired (P) and joining (J) regions (Figure *I*-1B). In addition, tertiary interactions have been predicted based on phylogeny and cross-linking experiments (Massire et al., 1998, Chen et al., 1998). M1 RNA of RNase P may be subdivided into two individual domains (D1 and D2; Figure *I*-1B), which are analogous to the two domains of the related *B. subtilis* RNA. The two domains of RNase P M1 RNA are the specificity domain (D1) responsible for ptRNA recognition and binding, and the catalytic domain (D2) containing the active site residues. Iron-EDTA footprinting experiments of Mg^{2+} dependent structure formation have suggested that the two domains of M1 RNA can fold independently (Loria and Pan, 1996; Pan et al., 1999). It has also been shown that separation of D1 and D2 in the *B. subtilis* RNA enables D2 to escape a kinetic trap in the

---

[1] Adapted from Kent, O.A., Chaulk, S.G., and MacMillan, A.M. (2000). *J. Mol. Biol.* **304**, 699-705.

overall folding of the RNA; separation of D2 from D1 speeds the overall folding of the former by ~15 fold while the folding rate of D1 remains unchanged (Loria and Pan, 1996; Fang et al., 1999; Pan et al., 1999).

The formation of tertiary folded structure in M1 RNA and the catalytic sub-unit of *B. subtilis* RNase P have been explored using oligonucleotide hybridization/RNase H cleavage, fluorescence, and CD spectroscopy (Zarrinkar et al., 1996; Fang et al., 1999). Oligonucleotide hybridization/RNase H cleavage was employed to examine the folding of RNase P RNAs from both *E. coli* and *B. subtilis* (Zarrinkar et al., 1996). These studies identified late events in the formation of the folded structure of both RNAs such as the formation of the P6 helical region (Figure *I*-1B). Fluorescence and CD studies of the folding of the domain analogous to M1 RNA D2 from *B. subtilis* RNase P RNA have shown that the two domains of RNase P RNA can fold independently. In addition, it has been suggested that discrete intermediates are involved in the formation of a final active structure although the exact nature of these intermediates remains unclear (Loria and Pan, 1996; Fang et al., 1999; Pan et al., 1999).

The creation of complex folded structures of RNA requires divalent magnesium and in many cases the folding process takes place over the course of several minutes. It has been proposed that the folding paths of large RNAs proceeds through discrete intermediates but the nature of these intermediates is not known in most cases. Here we describe our studies on the folding of the M1 RNA sub-unit of *E. coli* RNase P. We performed kinetic footprinting studies of M1 RNA folding with the chemical footprinting reagent peroxynitrous acid (King et al., 1992; King et al., 1993) to provide a detailed description of the folding pathway of RNase P RNA. In contrast to the Group I ribozyme, our results indicate that the M1 RNA folds into its catalytically active structure through two separately folded domains each containing a discrete series of intermediates. The formation of two independently folded domains may account for the observation of a kinetic trap in the folding of this RNA.

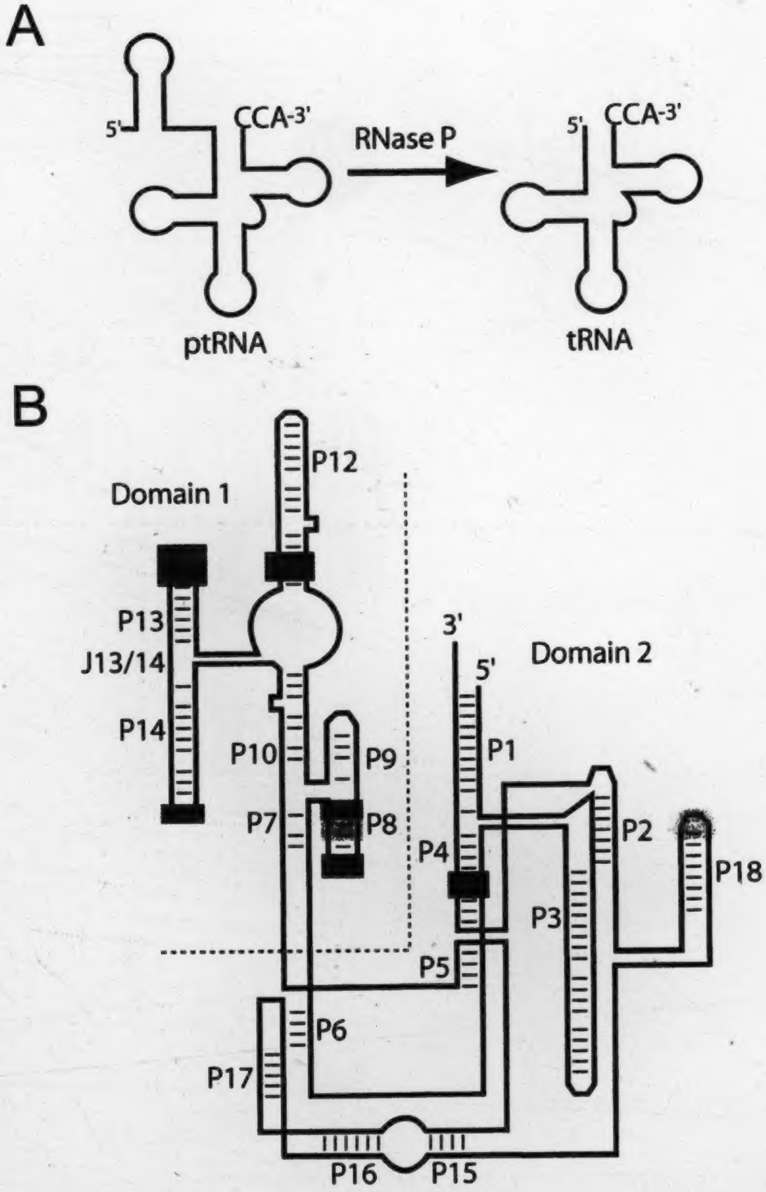


Figure I-1. RNase P. **(A)** RNase P cleavage of the 5' leader sequence from ptRNAs to generate the mature tRNA. The reaction is catalyzed by the holoenzyme, containing both protein (C5) and RNA (M1) components. **(B)** M1 RNA secondary structure showing the two domains and tertiary structural interactions (between like colored regions) proposed on the basis of phylogeny (after Massire et al., 1998).

I-2. Results and Discussion.

I-2.1. Endoribonuclease Activity and Folding of M1 RNA.

M1 RNA and the optimal RNase P substrate ptRNA^{Tyr} were transcribed as run-off transcriptions from the pJA2' and pTyr linearized plasmids respectively. The transcribed M1 RNA was tested for activity in the presence of standard RNase P high salt buffer and shown to be active by the appearance of the 5' leader sequence generated from the cleavage of 5'-end labeled ptRNA^{Tyr} (Figure *I*-2A). It was anticipated that the high salt conditions used for the activity assay would be incompatible with peroxynitrous acid cleavage. Therefore, we chose to use sodium phosphate buffer and 25 mM $MgCl_2$: conditions that had been used to footprint the Group I ribozyme with peroxynitrous acid (Chaulk and MacMillan, 2000a).

Standard conditions for M1 RNA folding and activity use 100 mM $MgCl_2$; however, 25 mM $MgCl_2$ can be used to fold the RNA but to maintain optimal activity the reaction conditions were substituted with potassium chloride. Therefore, reactions with low magnesium concentrations were doped with varying potassium chloride concentrations in the RNase P cleavage reaction. In order to determine the optimal amount of KCl in the reaction, we performed a KCl titration in the presence of 25 mM $MgCl_2$ and analyzed the M1 RNA activity by the appearance of the 5' leader sequence (Figure *I*-2B). KCl was titrated from 0-250 mM and optimal activity (assayed by comparison to standard conditions) was observed with 125 mM KCl. Therefore, these conditions were chosen for M1 RNA footprinting.

Before examining the folding pathway of the M1 RNA, the overall folding rate was determined using the endoribonuclease activity of M1 (Figure *I*-3A) as a measure of the amount of fully folded M1 present in a reaction (Pan et al., 1999). M1 RNA folding was initiated by the addition of magnesium chloride to a final concentration of 25 mM as described above. A ^{32}P-labeled M1 RNA substrate, ptRNA^{Tyr} from *E. coli*, was added after M1 RNA was allowed to fold for 0-1200 seconds. Following cleavage for twenty seconds, the reactions were quenched with NaOAc and ethanol and analyzed by denaturing PAGE (Figure *I*-3B).

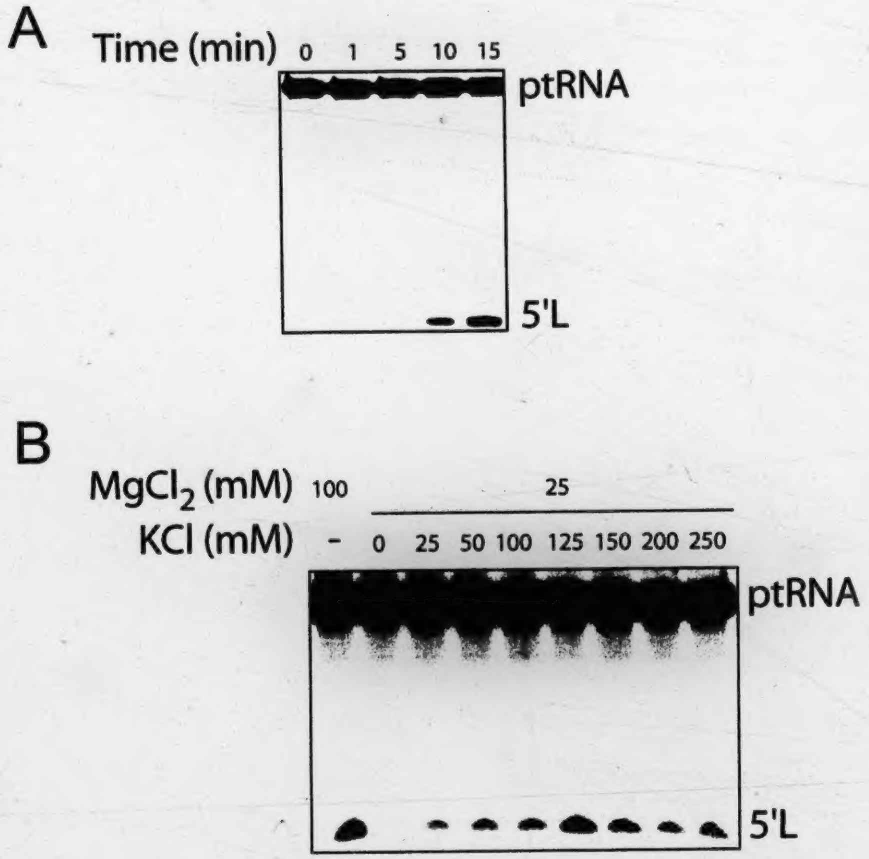


Figure I-2. **M1 RNA enzymatic activity assay.** **(A)** M1 RNA is catalytically active in the absence of protein component. The M1 RNA incubated with ptRNA$^{\mathrm{Tyr}}$ cleaves the 5' leader sequence (5'L) to give the mature tRNA$^{\mathrm{Tyr}}$. RNAs were incubated at 37°C for 0 to 15 minutes in the presence of 100 mM magnesium chloride. **(B)** M1 RNA activity assay. Potassium chloride titration in the M1 RNA activity assay in the presence of 25 mM magnesium chloride. Reactions were compared to the standard conditions (100 mM MgCl$_2$) to find the optimally substituted conditions. The ptRNA and 5' leader sequence (5'L) are indicated.

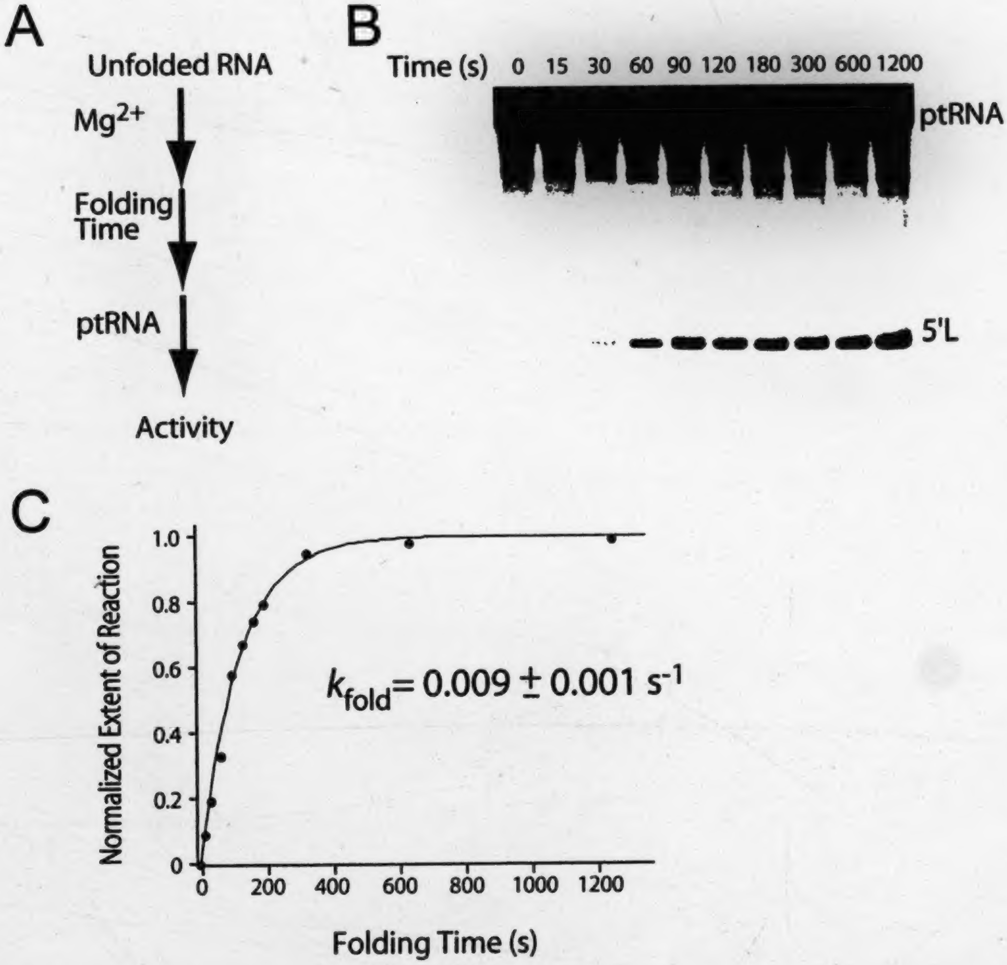


Figure I-3. Endoribonuclease assay of M1 folding. **(A)** Overview of the assay. The unfolded M1 RNA is incubated with $MgCl_2$ and allowed to fold, and then incubated with the pre-folded substrate RNA. Extent of reaction is a direct indication of the amount of correctly folded M1 RNA. **(B)** Endoribonuclease assay of Mg^{2+} dependent folding of the M1 RNA. M1 RNA was folded at 37°C for 0-1200 seconds and the extent of folding assayed by M1 enzymatic activity. The ptRNA and 5' leader sequence (5'L) are indicated. **(C)** The over all rate of folding was determined by measuring the catalytic activity of the RNA as a function of time following the addition of $MgCl_2$ to folding reaction. Folding time on the x-axis is the folding in the absence of substrate.

Under these conditions, the amount of RNA that folds during the reaction time will be negligible compared to the amount of RNA prefolded during the folding assay. The data from three independently run experiments was quantified and plotted (Figure *I*-3C). The half-time of RNA folding was determined to be ~1 minute; fitting this data to a first order rate equation gave a rate constant for folding, k_{fold}, of ~0.01 s^{-1}, which agrees with previously reported results for the rate constant of folding for the related *B. subtilus* RNA (Pan et al., 1999).

I-2.2. Kinetic footprinting of M1 RNA with peroxynitrous acid.

Upon folding of RNAs into their tertiary structures, buried regions are protected from hydroxyl radical induced cleavage; the protected or "footprinted" regions are indicative of the folded state of the RNA. We determined the equilibrium footprint of folded M1 RNA by subjecting both unfolded and folded RNA to cleavage with the hydroxyl radical source peroxynitrous acid (King et al., 1992; King et al., 1993; Chaulk and MacMillan, 2000) followed by a comparison of the cleavage patterns obtained in both cases. When these experiments were performed with 5' end-labeled RNAs, patterns of magnesium dependent protection consistent with the formation of a folded structure and consistent with the results of Fe-EDTA footprinting (Loria and Pan, 1996) were observed (Figure *I*-4A and Figure *I*-4B). The relative protections ranged from 50-90% and were found throughout the sequence of the M1 RNA.

In order to perform studies on the dynamics of M1 folding we made use of the kinetic footprinting properties of peroxynitrous acid (Chaulk and MacMillan, 2000). This reagent is unstable in aqueous solution at neutral pH and undergoes homolytic cleavage to produce hydroxyl radical and nitrogen dioxide. The decomposition process is rapid occurring with a half-life of ~1 second at pH 7. Thus, peroxynitrous acid is suitable for use in dynamic or "kinetic" footprinting studies and can be used to study processes with half-lives of ~3 seconds or greater. For example, peroxynitrous acid was previously utilized to analyze events in the folding pathway of the Group I ribozyme, which take place with half-lives ranging from 3-30 seconds (Chaulk and MacMillan, 2000).

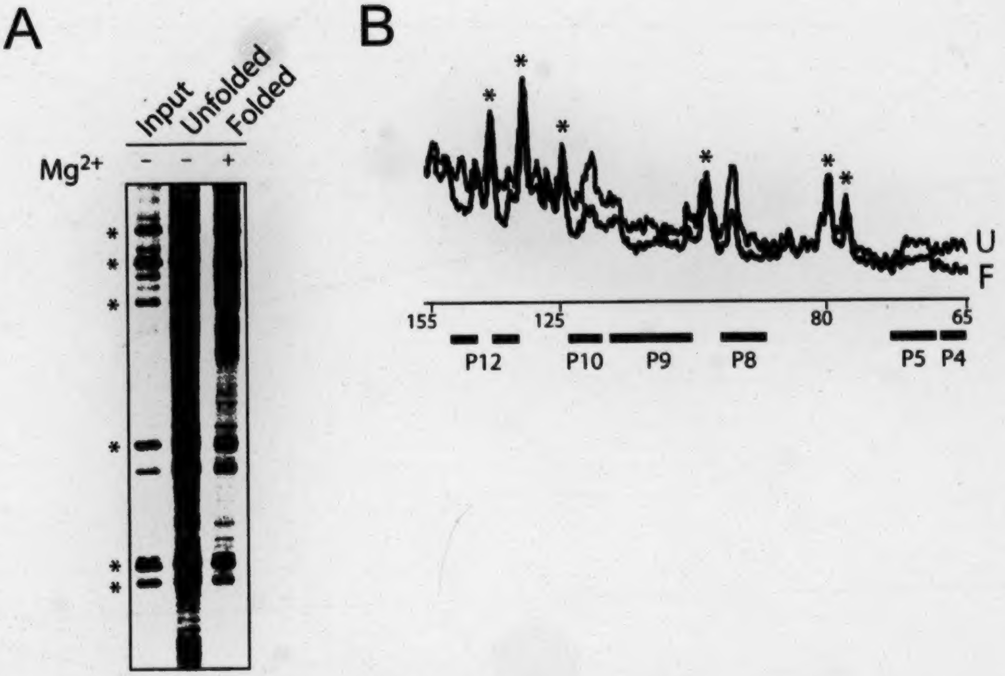


Figure I-4. Magnesium dependent protection of M1 RNA. **(A)** Analysis of footprinting unfolded and folded M1 RNAs. 5' end-labeled M1 RNAs were cleaved with peroxynitrous acid (Unfolded) and compared to M1 RNA prefolded in the presence of 25 mM Mg^{2+} (Folded). Both reactions are compared to the input RNA. Asterisks indicate degradation bands that appear at similar levels in all lanes. **(B)** Densitometric analysis of cleavage reactions. Blue; unfolded M1 RNA. Black; folded M1 RNA. Asterisks indicate the position of degradation bands that were used to normalize the graphs. Nucleotide positions indicated. Regions of Mg^{2+} dependent protection indicated with solid black lines referenced to regions of M1 RNA.

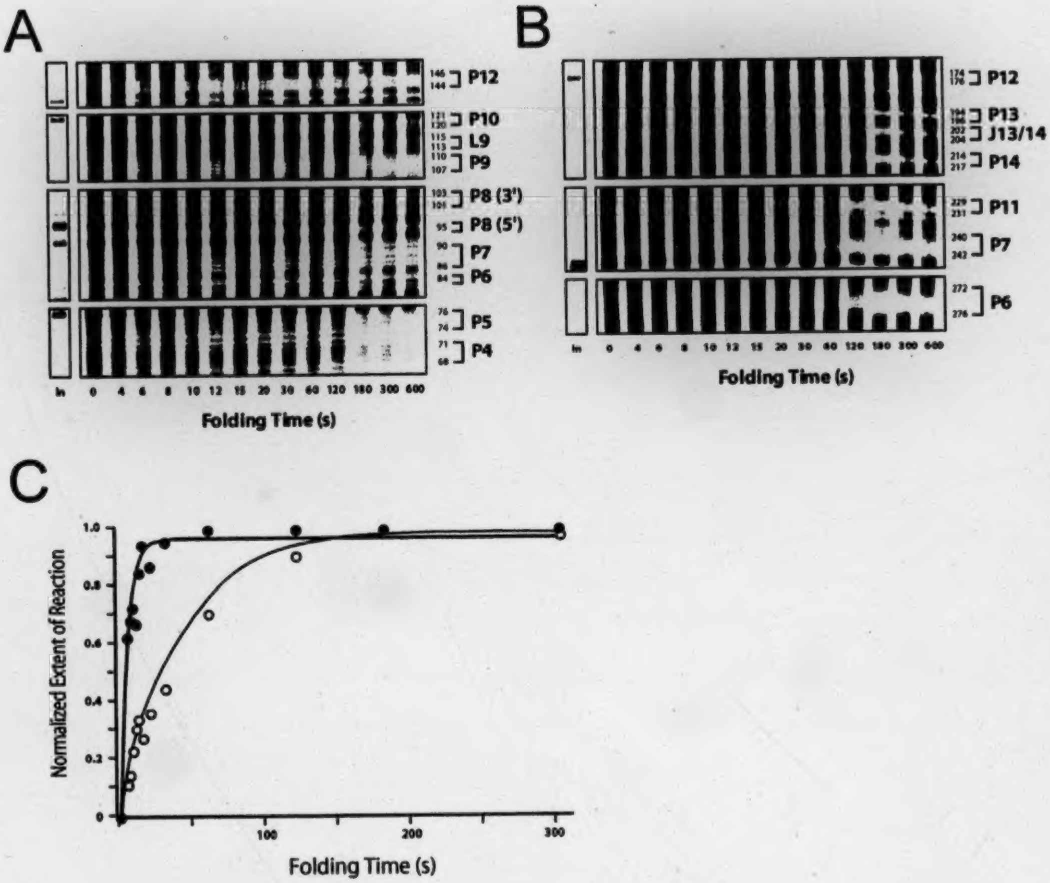


Figure I-5. Kinetic footprinting of the M1 RNA folding pathway using peroxynitrous acid. **(A)** Denaturing PAGE analysis of 5' end-labeled RNA treated with peroxynitrous acid after folding for the indicated times following addition of 25 mM magnesium to the folding reaction. Peroxynitrous acid footprinted RNAs were compared to input RNA (In). Nucleotides protected and their corresponding regions in M1 RNA indicated. **(B)** Denaturing PAGE analysis of 3' end-labeled RNA treated with peroxynitrous acid at the indicated times following addition of magnesium to the folding reaction. **(C)** Representative relative protections of fast (P12; ●) and slow (P6; ○) folding regions of the M1 RNA (600 and 900 second time points are not shown).

The potassium salt of peroxynitrous acid is stable in basic solution and therefore provides a convenient means of storing the reagent. Stock solutions of potassium peroxynitrite were prepared by oxidation of nitrous acid followed by basification with potassium hydroxide (King et al., 1992; King et al., 1993; Chaulk and MacMillan, 2000). Hydroxyl radical formation could then be initiated by simple addition of this precursor to reactions buffered at pH 7.

Kinetic footprinting studies were conducted with both 5' and 3' end-labeled M1 RNA. Folding of labeled M1 RNA was initiated by adding divalent magnesium to folding reactions followed by the addition of potassium peroxynitrite after various folding times. Following cleavage, reactions were quenched by incubation on dry ice. Reactions were desalted by precipitation and analyzed by denaturing PAGE (Figure *I*-5A). As was observed in equilibrium footprinting experiments, there were regions of significant protection observed upon folding of the M1 RNA. In addition, it was clear that the amount of protection was time dependent and varied in different regions of the RNA. In order to assess the precise time and domain dependence of such protections, cleavages were quantified over specific regions and plotted as a fraction of relative protection versus time (Figure *I*-5C; data not shown; Chaulk and MacMillan, 2000). Assuming a first order process, the data was fit to a first order exponential to yield rates for individual protections (Figure *I*-5C and Table *I*-1).

Similarly, we prepared 3' end-labeled M1 RNA and carried out kinetic footprinting experiments of the folding pathway using peroxynitrous acid as described above (Figure *I*-5B). Once again, a number of time dependent protections were observed. We quantified these protections as a function of time and fit the data to a first order exponential (Figure *I*-5C) to give rate constants for the individual protections (Table *I*-1).

I-2.3. Implications of kinetic analysis on M1 RNA folding and structure.

The protections observed in this study occurred with rate constants ranging from 0.20-0.02 s^{-1} that correspond to half-lives of protection ranging from 4 to 35 seconds (Table *I*-1). Seventeen separate kinetic footprints were identified spanning most of the M1 sequence (except for the extreme 5' and 3' ends of the RNA; Figure *I*-6A and Table *I*-

1). Kinetics of protection of the P6 helix are in good agreement with data obtained from oligonucleotide hybridization/RNase H cleavage which was used to examine a late event in M1 folding (Zarrinkar et al., 1996). The protections observed in the present study fall into several classes and may be broadly categorized as fast, medium, and slow events (Figure *I*-6). These observations have implications for both the folding path of the M1 RNA as well as the nature of the final tertiary structure.

Table I-1. Rates of folding for regions of M1 RNA observed using peroxynitrous acid as a kinetic footprinting reagent.

Region [a,b]	*Nucleotides Protected*	Rate (s^{-1})[c]
P4[a]	68-71	0.06
P5[a]	74-76	0.07
P6(5')[a]	84	0.04
P7[a]	86-90	0.06
P8(5')[a]	95	0.07
P8(3')[a]	101-103	0.06
P9[a]	108-110	0.06
L9[a]	113-115	0.05
P10[a]	120-121	0.07
P12[a]	144-146	0.17
P12[b]	174-176	0.16
P13[b]	194-196	0.20
J13/14[b]	202-204	0.03
P14[b]	214-217	0.02
P11[b]	229-231	0.05
P7[b]	240-242	0.04
P6[b]	272-276	0.02

[a] 5' end-labeled M1 RNA
[b] 3' end-labeled M1 RNA
[c] Uncertainties in rates are one standard deviation from the calculated value.

The earliest folding events observed here involve the P12 and P13 helices of D1 in regions proposed to interact via tertiary interactions (Massire et al., 1998). The rapid formation of structure in this region is reminiscent of the formation of the P4/P6 core of the Group I ribozyme. However, in contrast to the Group I folding pathway, it is unlikely that these early folding events serve as a scaffold for folding events in a separate domain since it has been demonstrated that D2 can form in the absence of D1 (Loria and Pan, 1996; Pan et al., 1999).

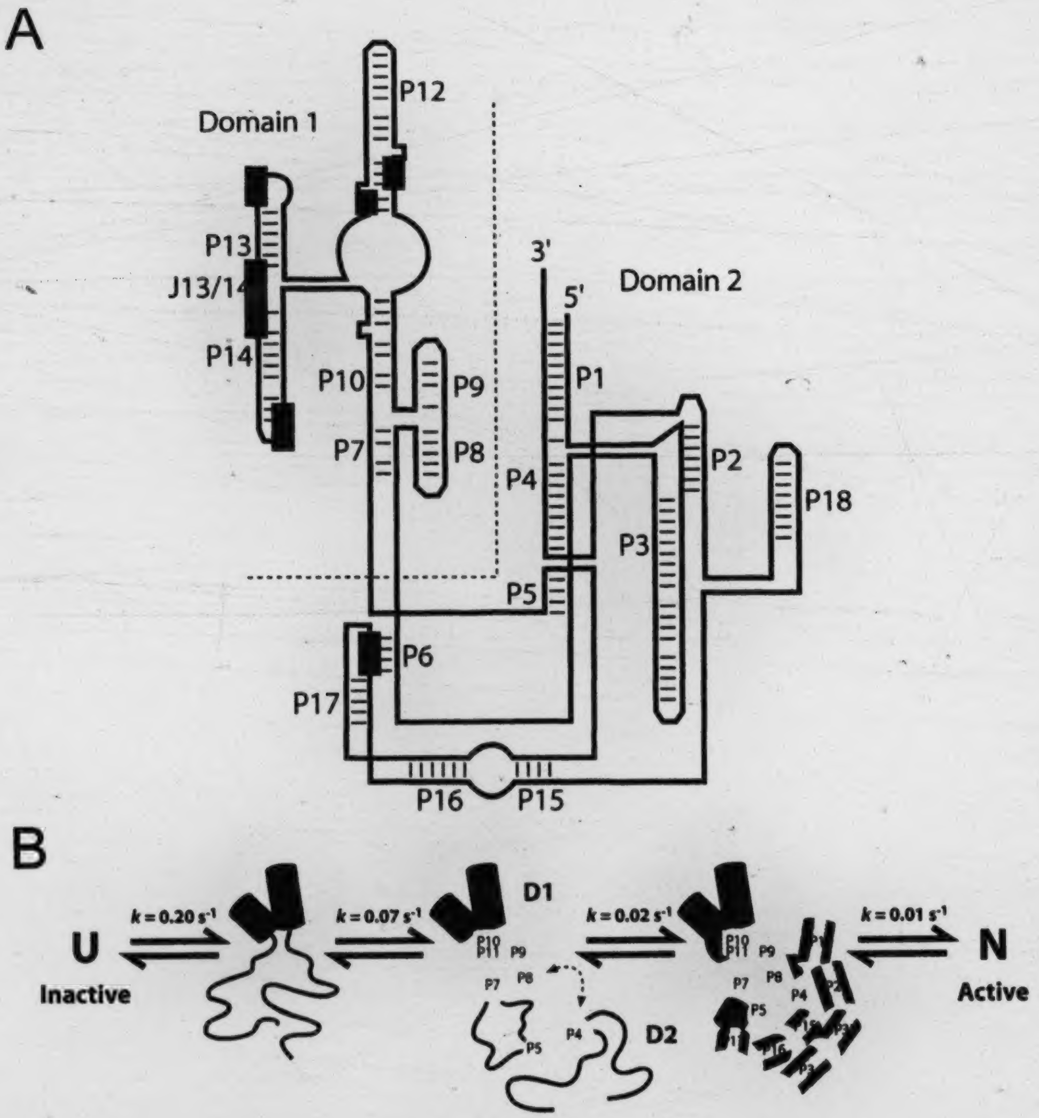


Figure I-6. Folding pathway of RNase P M1 RNA. **(A)** Secondary structure (adapted from Massire et al., 1998) of the M1 RNA consistent with predicted tertiary structure showing the hierarchy of protections observed due to the formation of RNA structure during the folding experiment: fast (green), medium (yellow), and slow (red). **(B)** Summary of events in the folding pathway of M1 RNA. Stepwise formation of the active tertiary structure of M1 RNA from the unfolded (U) to the native (N) states showing fast (green), medium (yellow), and slow (red) events (after the model of Massire et al., 1998). Double-headed arrow represents possible inter-domain interactions during M1 folding.

Intermediate rates of folding (half-lives of 10-17 seconds) were observed for regions of both D1 and D2 (Table *I*-1). Protection of the P7 through P11 helices represents the formation of the essential elements of D1 while protection of P4, P5, and the 5' side of P6 correspond to structure formation in D2 (Figure *I*-6). Thus, in contrast to the Group I RNA, the formation of the two domains in M1 can be described as parallel sets of events. Kinetic footprinting with hydroxyl radicals generated by synchrotron X-ray irradiation has permitted the elucidation of the detailed stepwise folding pathway of the Group I ribozyme (Sclavi et al., 1997; Sclavi et al., 1998), which folds into its active structure through a nucleating core.

In addition, the high resolution X-ray structure of a large fragment of the Group I RNA has also been determined (Golden et al., 1998) and consists of the core P4/P6 domain which stabilizes a set of helices that are essentially wrapped around it; the observation that P4/P6 folds rapidly with respect to the rest of the RNA thus accords well with this structure. However, the fact that both D1 and D2 of M1 RNA fold with similar rates suggests that there is not a nucleating core for the overall folding process. It is intriguing that intermediate steps in the folding of M1 RNA occur at the interface of the D1 and D2 domains. This is consistent with the existence of a kinetic trap in the folding process involving interactions between D1 and D2 (Fang et al., 1999; Pan et al., 1999).

Three additional sets of protections are observed in the kinetic folding study: the 3' side of P6, the P14 helix and J13/14 region are protected with half-lives of ~23-35 seconds (Table I-1). P6 is part of D2 while both P14 and J13/14 reside in D1. P6 is formed from nucleotides far apart in the linear sequence of D2 and likely serves to hold D2 in a tertiary fold. The protection observed is probably due to its interactions with P7 of D1 and P5 of D2 (Chen, et al., 1998; Massire et al., 1998). The protection observed in P14 is likely due to its tertiary interaction with P8 of the 4-helix junction while protection of J13/14 may be due to its interaction with the P10-P12 helices (Chen et al., 1998; Massire et al., 1998). The P14 and J13/14 protections represent a late rearrangement of tertiary structure within D1, whereas the protection of P6 represents a possible inter-domain interaction between P6 and P7 of D1 (Chen et al., 1998; Massire et al., 1998). Both of these protections occur late on the pathway to the final active folded structure

(Figure *I*-6B). These protections are possibly associated with escape from an earlier kinetic trap involving the 3' side of P6 (Fang et al., 1999; Pan et al., 1999).

Interestingly, none of the protections observed in the present study correspond to the overall rate of folding as measured by the endonuclease activity assay (Figure *I*-3 and Table *I*-1). Although this difference may be due to differences between the two folding assays, it is consistent with subtle rearrangements that may be required for the formation of a fully active structure. Separation of the two domains in *B. subtilis* results in fast folding of D2 but no change in the rate of folding of D1 (Pan et al., 1999); the magnitude of rates observed in the present study suggests that a kinetic trap in the folding of M1 RNA is likely to be an early event in the folding pathway.

This study suggests an ordered, parallel pathway for individual folding of the two domains of M1 RNA, followed by interaction of these two domains into the biologically active form. In this regard, M1 folding significantly differs compared to the only other large RNA folding path which has been studied in detail, that of the Group I ribozyme. In further contrast to the Group I RNA, the folding of the separate domains of M1 is a concerted process that almost certainly involves inter-domainal interactions and one or more kinetic traps. Thus, although an ordered pathway of folding for large RNAs may be general, the detailed mechanism of folding is unique to a specific RNA.

I-3. Materials and Methods.

I-3.1. Synthesis of M1 RNA and ptRNA substrates.

Plasmid purification. The pJA2' plasmid containing the RNase P M1 RNA and pTyr plasmid containing an optimal RNase P ptRNA^{Tyr} RNA both containing an ampicillin resistance cassette were transformed into competent *E.coli* BL21 cells by incubation on ice followed by incubation in LB at 37°C for 1 hour. Following transformation, cells were plated onto LB-agar plates containing 50 µg/mL ampicillin and grown overnight at 37°C. Isolated colonies were picked and grown in LB-amp broth overnight at 37°C. Overnights were centrifuged to pellet the cells, the supernatant discarded and the cell pellet resuspended into lysis buffer (Qiagen). Plasmid DNA was purified from cellular debris and bacterial DNA by mini-prep kit (Qiagen) following the manufacturers protocol. Isolated purified plasmids were ethanol precipitated and resuspended in distilled water. Plasmid concentration was determined using a UV spectrophotometer at an absorbance of 260 nm.

Digestion of plasmid DNA. The pJA2' and pTyr plasmids were engineered to contain FokI restriction sites at the 3' end of the encoded RNA sequence. Digestion of the plasmid DNA with Fok1 facilitated run-off transcriptions of the encoded RNAs. Digestion reactions contained 15 ug plasmid DNA, 1×NEB4 buffer, and 40 units FokI restriction enzyme. Reactions were incubated at 37°C for 2 hours, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. An aliquot of the reactions was analyzed on 1.5 % (w/v) agarose gels with 1×TBE running buffer at 85 mA for 45 minutes. Comparing uncut to cut plasmids demonstrated that reactions proceeded to completion.

Transcription of M1 and ptRNA^{Tyr} RNAs. RNAs were transcribed from the digested plasmid DNA. Run-off transcription reactions contained 40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 0.5 mM NTPs, 10 mM DTT, 100-400 ng cut plasmid DNA template, and 10 units RNA polymerase. Transcriptions were incubated at 37°C for 4 hours, centrifuged to remove insoluble inorganic phosphate accumulated during the reaction, extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol and ethanol precipitated. In order to create M1 RNA with homogenous 3' ends, a low magnesium transcription procedure was employed. Low magnesium transcriptions contained 40 mM Tris-HCl, pH 7.9, 3 mM $MgCl_2$, 2 mM spermidine, 25 mM NaCl, 1 mM NTPs, 4 mM DTT, 100-400 ng cut plasmid DNA template, and 8 units RNA polymerase. Transcriptions were incubated at 37°C for 1.5 hours, and precipitated as described above. RNAs were resuspended into loading dye/water (1:3 v/v) and loaded onto 8% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. RNAs were visualized by UV shadowing, excised from the gel, and purified as described previously.

Dephosphorylation and 5'/3' end labeling. Transcribed RNAs were 5' dephosphorylated using calf intestinal alkaline phosphatase (CIAP, New England Biolabs). CIAP reactions contained 100-500 pmol RNA, 1×NEB3 buffer, 0.5 units RNase Inhibitor (Roche), and 50 units CIAP. Reactions were incubated at 37°C for 30 minutes, made 0.3 M in sodium acetate, extracted with phenol/chloroform/isoamyl alcohol twice, chloroform/isoamyl alcohol and ethanol precipitated. Precipitated RNAs were resuspended in 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 50 μCi γ-[^{32}P]-ATP, and 5 units T4 polynucleotide kinase (Invitrogen). Labeling reactions were incubated at 37°C for 30 minutes. 3'-end labeling reactions contained 20 mM Tris-HCl, pH 7, 50 mM KCl, 0.7 mM $MnCl_2$, 0.2 mM EDTA, 100 μg/mL acetylated BSA, 10% (v/v) glycerol, 0.6 μM M1 RNA, 125 μCi α-[^{32}P]—cordycepin, and 1500 units yeast Poly(A) polymerase (Amersham Life Science). Reactions were incubated at 30°C for 20 minutes. Following labeling reactions, RNAs were loaded directly onto 6% (19:1 acrylamide/bisacrylamide) denaturing gels and purified as described above.

I-3.2. M1 RNA activity and folding.

M1 RNA activity assay. M1 RNA was pre-folded in 50 mM Tris-HCl, pH 8, and 100 mM $MgCl_2$, heated to 55°C for 5 minutes then slow cooled to room temperature, and then transferred to ice. Pre-tRNA^{Tyr} was pre-folded in 50 mM Tris-HCl, pH 8, 100 mM $MgCl_2$, and 1 M KCl, heated to 98°C for 2 minutes, and then cooled as described above. Folded RNAs were incubated together in reactions that contained 2 μM M1 RNA and 1 nM ptRNA^{Tyr} (50-100×10^3 c.p.m.) at 37°C for 5-60 minutes, made 0.3 M in sodium acetate and ethanol precipitated. Pelleted RNAs were resuspended in loading dye and loaded onto 15% (19:1 acrylamide/bisacrylamide) denaturing 8 M urea gels. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Salt titration. In order to footprint the M1 RNA under footprinting conditions used to analyze Group I ribozyme folding (Chaulk and MacMillan, 2000a) the optimal activity of

M1 RNA in phosphate buffer was determined. Reactions containing 20 nM M1 RNA, 5 nM ptRNA^{Tyr} (50-100×10^3 c.p.m.), 50 mM NaH_2PO_4/Na_2HPO_4, pH 7, and 25 mM $MgCl_2$ were titrated with 0-250 mM KCl, and incubated at 37°C for 10 minutes. Reactions were quenched with loading dye (6.4 M urea, 0.4 mg/mL BPB/XC, 0.4 M Tris-base, and 0.4 M boric acid) and subjected to 15% (19:1 acrylamide/bisacrylamide) denaturing PAGE. Gels were exposed to a Molecular Dynamics phosphor screen and scanned using a Molecular Dynamics Storm 840 Phosphorimager.

Folding assay. The overall rate of RNA folding was determined by measuring the catalytic activity of the RNA as a function of time following the addition of Mg^{2+} to folding reactions. Aliquots (10 μL) from a ribozyme stock solution (20 nM ribozyme, 120 mM KCl, 25 mM $MgCl_2$, 50 mM NaH_2PO_4/Na_2HPO_4, pH 7) were removed after folding for 0-1200 seconds following $MgCl_2$ addition, and then combined with aliquots (10 μL, 1×10^4 c.p.m.) of a substrate stock solution (5 nM 5'-end labeled ptRNA, 500 mM KCl, 100 mM $MgCl_2$, 50 mM NaH_2PO_4/Na_2HPO_4, pH 7). Reactions proceeded for 20 seconds before quenching with an equal volume of stop solution (100 mM Na_2EDTA, 6.4 M urea, 0.4 mg/mL BPB/XC, 0.4 M Tris-base, and 0.4 M boric acid). Reactions were then subjected to 15% (19:1 acrylamide/bisacrylamide) denaturing PAGE for 1.5 hours at 75 W. Gels were exposed to a Molecular Dynamics Phosphor Screen that was then scanned on a Molecular Dynamics Storm 860 Phosphorimager. Extents of reaction were quantitated (ImageQuant 5.0) and a plot of normalized extent of reaction versus folding time was fitted to a first order exponential (GraFit 3.00) using data from three experiments.

I-3.3. Kinetic footprinting M1 RNA.

Synthesis of potassium peroxynitrite. An aqueous solution containing 0.6 M $NaNO_2$ and 3% (v/v) H_2O_2 was stirred on ice until cold. A cold solution of 1.2 M HCl was added and immediately chased with cold 1.6 M KOH and 400 uM diethylene-triamine-pentaactetic acid. The solution was stirred for 5 minutes on ice, and then 100 mg MnO_2 was added and stirring continued for 20 minutes. The excess MnO_2 was removed by centrifugation at 4°C. Typical yields of potassium peroxynitrite were 80-130 mM by UV absorbance at 302 nM (ε=1670 M^{-1}).

Footprinting M1 RNA. RNA for footprinting studies was 5'-end-labeled or 3' end-labeled as described above. M1 RNA folding was initiated by the addition of $MgCl_2$ (final concentration 25 mM) to a solution (final volume 20 μL) containing final concentrations of 125 mM KCl, 50 mM NaH_2PO_4/Na_2HPO_4, pH 7, 5 nM 5' or 3' end-labeled ribozyme ($0.5\text{-}3.0\times10^6$ c.p.m.) at 37°C. Footprinting was then effected by the addition of peroxynitrous acid (1 μL, 80-130 mM) after 0 to 600 seconds of folding. Footprinting reactions were allowed to proceed for 20 seconds before freezing on dry ice followed by ethanol precipitation. Reactions ($0.5\text{-}3.0\times10^6$ c.p.m.) were then subjected to 6% (19:1 acrylamide/bisacrylamide) denaturing PAGE at 85 W for 2 hours for 5' end-labeled RNA and 1.5 hours for 3' end-labeled RNA. Dried gels were exposed to a Molecular Dynamics Phosphor Screen that was then scanned on a Molecular Dynamics Storm 860 Phosphorimager. To calculate the fractional peroxynitrous acid protection (footprint), background intensity (ImageQuant 5.0) was first subtracted and a correction for differences in lane loading was applied. Then the intensity of a region at a given time was divided by the intensity of the same region at time zero. The resulting fractional intensities were normalized from 0 to 1. Plots of normalized fractional protection versus folding time were fitted to a first order exponential (GraFit 3.00) using data averaged from three separate experiments.

I-4. References.

Altman, S. (1999). Ribonuclease P. In *The RNA World, 2nd Edition* (Gesteland, R.F., Cech, T.R. & Atkins, J.F.) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Batey, R.A., Rambo, R.P., and Doudna, J.A. (1999). Tertiary Motifs in RNA Structure and Folding. *Agnew. Chem. Int. Ed.* **38**, 2326-2343.

Chaulk, S.G., and MacMillan, A.M. (2000a). Characterization of the Tetrahymena ribozyme folding pathway using the kinetic footprinting reagent peroxynitrous acid. *Biochemistry* **39**, 2-8.

Chaulk, S.G., and MacMillan, A.M. (2000b). Kinetic Footprinting of an RNA-Folding Pathway Using Peroxynitrous Acid. *Angew. Chem. Int. Ed.* **39**, 521-523.

Chen, J.L., Nolan, J.M., Harris, M.E., and Pace, N.R. (1998). Comparative photocross-linking analysis of the tertiary structures of Escherichia coli and Bacillus subtilis RNase P RNAs. *EMBO J.* **17**, 1515-1525.

Fang, X., Pan, T., and Sosnick, T.R. (1999). Mg^{2+}-dependent folding of a large ribozyme without kinetic traps. *Nature Struct. Biol.* **6**, 1091-1095.

Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. (1998). A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259-264.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N.R., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857.

James, B.D., Olsen, G.J., Liu, J., and Pace, N.R. (1988). The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* **52**, 19-26.

King, P.A., Anderson, V.E., Edwards, J.O., Gustafson, G., Plumb, R.C., and Suggs, J.W. (1992). A stable solid that generates hydroxyl radical upon dissolution in aqueous solutions: reaction with proteins and nucleic acid. *J. Am. Chem. Soc.* **114**, 5430-5432.

King, P.A., Jamison, E., Strahs, D., Anderson, V.E., and Brenowitz, M. (1993). 'Footprinting' proteins on DNA with peroxonitrous acid. *Nucleic Acids Res.* **21**, 2473-2478.

Loria, A., and Pan, T. (1996). Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* **2**, 551-563.

Massire, C., Jaeger, L., and Westhof, E. (1998). Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773-793.

Pan, T., Fang, X., and Sosnick, T. (1999). Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *J. Mol. Biol.* **286**, 721-731.

Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S.A. (1998). RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279**, 1940-1943.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R., and Brenowitz, M. (1997). Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J. Mol.Biol.* **266**, 144-159.

Westhof, E., and Altman, S. (1994). Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5133-5137.

Zarrinkar, P.P., Wang, J., and Williamson, J.R. (1996). Slow folding kinetics of RNase P RNA. *RNA* **2**, 564-573.

Appendix *II*.
Synthesis of a convertible cytosine ribonucleotide.

Appendix *II*. Synthesis of a convertible cytosine ribonucleotide.

II-1. Introduction.

II-1.1. Modified nucleosides.

The use of modified nucleosides has allowed for the site specific attachment of reporter groups to RNA in order to probe both the global conformation of RNA molecules and the features of RNA-protein interactions (Allerson, 1996; Allerson and Verdine, 1995). For example, the attachment of fluorophore molecules at the 5' and 3' hydroxyl groups of an RNA molecule has been used to analyze ribozyme kinetics and protein RNA interactions by fluorescence resonance energy transfer (FRET) (Ghetu et al., 2002; Singh, et al., 1999; Walter et al., 1998). In addition, many chemical nucleases, such as 1,10-phenanthroline or EDTA derivatives, have been covalently attached to RNA via the 5' hydroxyl (Helene, 1998; Cech and Wang, 1992). Unfortunately, most RNA molecules are quite large with very complex tertiary structures that therefore decrease the value of studying these molecules at one end. Ideally, the incorporation of a reporter group to a specific central position would be preferred.

Although transcription reactions with T7 RNA polymerase in the presence of modified natural bases result in the incorporation of the modified bases into the RNA chain (Dervan and Yitzhak, 1993), the enzymatic procedure does not allow for a single site-specific modification into an RNA molecule. A number of chemically modified nucleosides have been incorporated into RNA molecules using solid phase RNA synthesis technology (Allerson, 1996). For example, 2'-OMe nucleosides, arabinonucleosides, 2'-amino nucleosides and 2'-fluoronucleosides incorporate modifications on the ribose sugar moiety of the RNA. Base modified nucleotides have also been developed employing covalently attached linkers or modification to various positions of the purine and pyrimidine heterocycles (Rana et al., 1999). Specifically, fully protected 4-thiouridine can be incorporated via automated DNA/RNA synthesis into an oligonucleotide to allow attachment of various reporter groups (Eshaghpour et al., 1979; Qin et al., 2003).

II-1.2. The convertible nucleotide strategy.

The convertible nucleotide strategy allows DNA or RNA sequences to be synthesized that contain a site specifically functionalized nucleoside (MacMillan and Verdine, 1991). The convertible nucleotide strategy utilizes the incorporation, via solid phase synthesis, of nucleotide derivatives that possess a leaving group on the nucleotide base at the position of the exocyclic amine. Following incorporation during automated synthesis, the resulting oligonucleotide can be treated with a nucleophile tethered to a functional group that will react specifically with the leaving group on the convertible nucleoside (Figure *II*-1A). Displacement of the leaving group and incorporation of the nucleophile-functional group will generate N-functionally tethered oligonucleotides (MacMillan and Verdine, 1991). The idea of a convertible nucleotide approach is that the leaving group on a modified monomer can be displaced after the DNA/RNA sequence has been synthesized.

The use of the convertible nucleoside in automated DNA/RNA synthesis yields a convertible oligonucleotide, which can be reacted with bifunctional nucleophiles. One end of the bifunctional nucleophile becomes attached to the DNA/RNA and the other end is available in solution for further reaction (MacMillan and Verdine, 1991). There are a number of criteria that are critical to the success of the convertible nucleotide approach: (i) the convertible nucleotide must be compatible with automated DNA/RNA synthesis procedures; (ii) it must undergo clean and rapid conversion (less than 24 hours) under conditions tolerated by RNA; and (iii) the convertible nucleotide must direct the functional tether to a non-sterically demanding region of RNA and permit placement within duplex regions (MacMillan and Verdine, 1991). In addition, the modification should not involve the 5' or 3' positions, since the attachment of the tether at these positions would prevent desirable enzymatic transformations such as end labeling and ligation.

The RNA structural implications of modification are that the tether should be flexible so as not to prevent Watson-Crick base pairing functionality (Allerson, 1996). The incorporation of alkyl amines into nucleic acids has been determined by structural modeling to take place in the major groove of the nucleic acid duplex (MacMillan and Verdine, 1991).

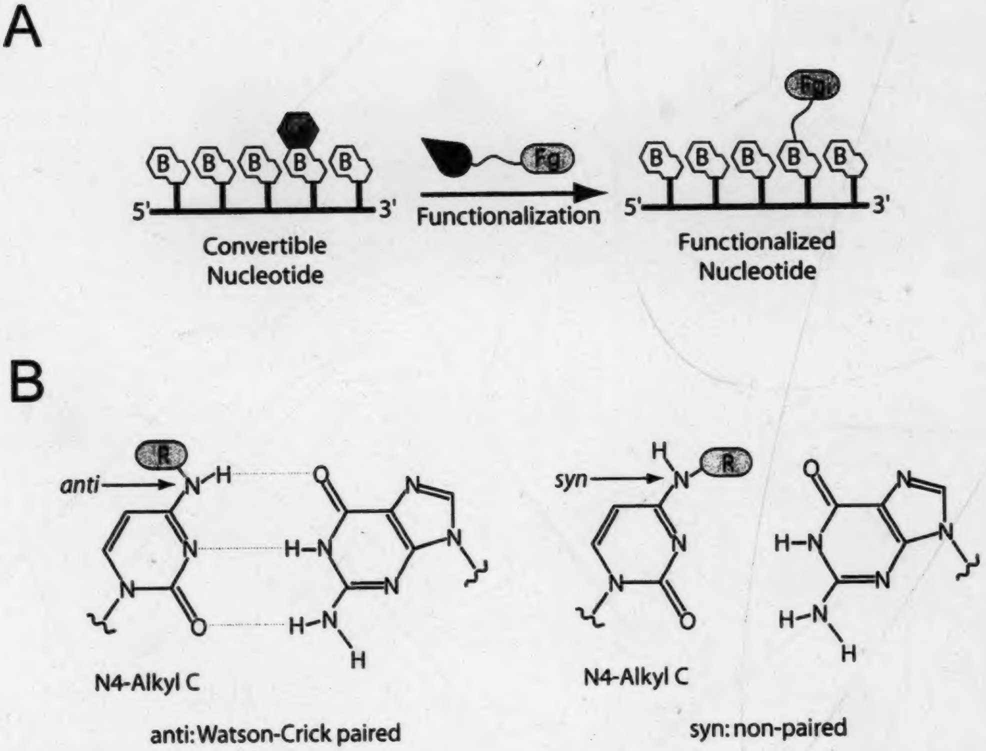


Figure II-1. The convertible nucleotide strategy. **(A)** From the RNA oligonucleotide, a modified base containing a leaving group (Lg) can be reacted with a nucleophile (Nu) containing a tethered functional group (Fg). Upon reaction the leaving group is displaced and the functional group is tethered to the RNA oligonucleotide. **(B)** Anti versus syn orientation of the alkyl tether. The anti rotomer of N4-alkyl cytosine (left) is thermodynamically disfavored in the free nucleoside, but favored in duplex oligonucleotides because it can base pair with guanine, unlike the syn rotomer (right) which disrupts Watson-Crick base pairing.

In the free nucleoside, the *syn* rotamer is favored; however, in the base paired nucleic acid the thermodynamically unfavorable *anti* rotamer of the free nucleoside is assumed since it permits base pairing (Figure *II*-1B). In contrast, the *syn* rotamer does not permit Watson-Crick base pairing (MacMillan and Verdine, 1991).

II-2. Results and Discussion.

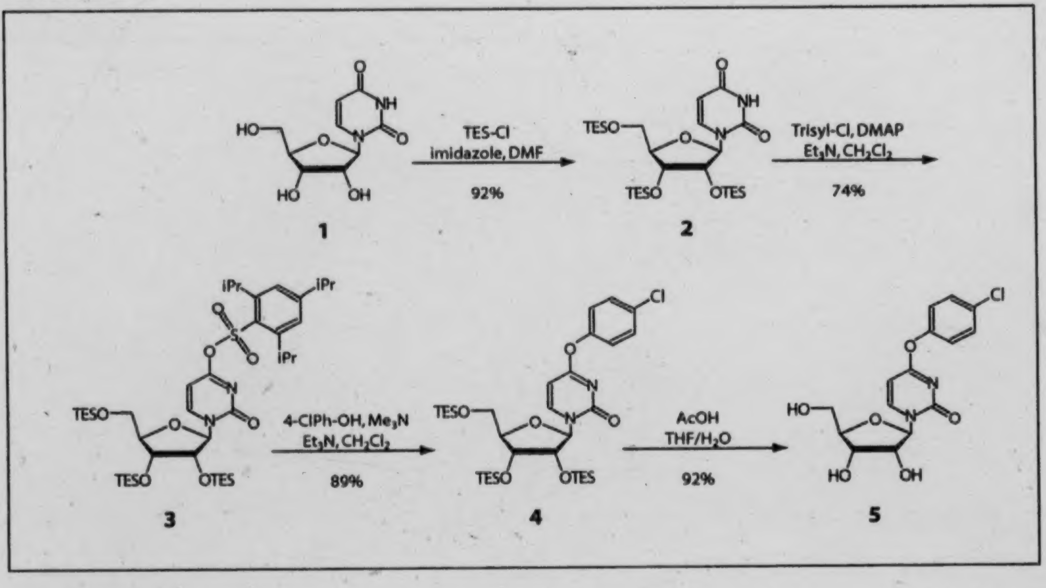
II-2.1. Synthesis of 4-chlorophenyl uridine.

It was decided from the work of Allerson (Allerson and Verdine, 1995), that 4-chlorophenyl-uridine (ClϕU) would be used as the convertible nucleoside in the RNA to be studied. The modification used by Allerson on the C4 position of uridine allows the eventually incorporated alkyl amine tether to be positioned in the major groove. Allerson previously investigated the chlorophenyl moiety as a good leaving group to be used in RNA convertible nucleotides (Allerson, 1996). Specifically, it was found that chlorophenyl-uridine ethers were sufficiently reactive with amine nucleophiles such that deprotection and functionalization could be attained during the same step. The convertible nucleoside ClϕU was synthesized as outlined in Scheme *II*-1.

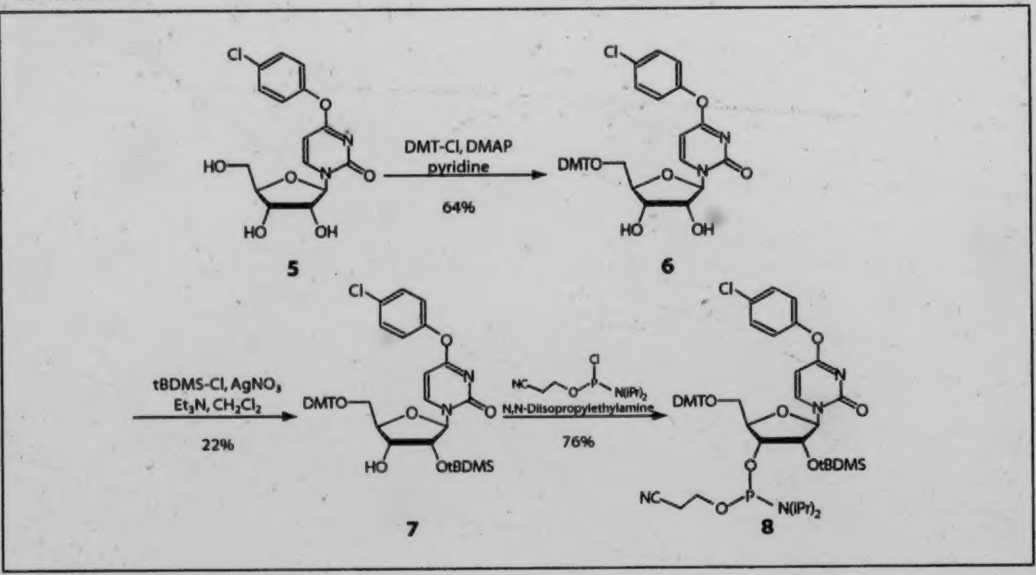
Starting from uridine **1** the ribose hydroxyls were protected by triethylsilane (TES) moiety to obtain the fully protected *tris*-TES nucleoside **2**. The TES ethers were easily removed under acidic conditions; conditions that were compatible with the presence of the aryl ether leaving group. The formation of the imide sulfonate ester **3**, followed by substitution with 4-chlorophenol led to the *tris*-TES-uridine aryl ether **4**. The TES groups were hydrolyzed with acetic acid in a water/THF mixture to yield the convertible nucleoside 4-O-(4-chlorophenyl) uridine **5** (ClϕU). Unfortunately, the TES deprotection reaction typically proceeded with poor yields, which is in contrast to those reported elsewhere (Allerson, 1996).

The convertible nucleotide ClϕU was then prepared as a fully protected phosphoramidite to be used in automated solid phase RNA synthesis (Scheme *II*-2). The 5'-OH was blocked with the acid labile 4,4'-dimethoxytrityl (DMT) group, the 2'-OH was protected as a *t*-butyldimethylsilyl (tBDMS) ether, and the 3'-OH was phosphitylated to generate an N,N-diisopropyl-phophoramidite functionality for coupling to the growing oligonucleotide.

Scheme II-1.



Scheme II-2.



Blocking of the 5'-OH of **5** with DMT proceeded cleanly with good yield to produce **6**. The preparation of the 2'-O-tBDMS intermediate **7** was a disappointing reaction. The best that can usually be achieved is a 1:1 mixture of the 2'-O-tBDMS and 3'-O-tBDMS products (Allerson, 1996). However, typical purified yields were in the 18 - 22% range of the desired product. Activation of the 3'-OH with the phosphoramidite moiety went smoothly with good yield using the necessary conditions, to give the fully protected convertible phosphoramidite **8**.

II-2.2. Synthesis of a functionalized RNA.

Deprotection of ribo-adenosine (rA) and ribo-guanosine (rG) residues is cleanly obtained using alkylamines. However, previous investigations had demonstrated that deoxycytidine residues protected with standard N4-benzoyl group during solid phase synthesis could undergo a side reaction when treated with alkylamines during deprotection (MacMillian and Verdine, 1991). The side reaction is caused by nucleophilic attack of the amine nitrogen at the C-4 position rather than the benzoyl carbonyl, which results in misincorporation of N4-cytosine tethers (Figure *II*-2). Fortunately, the side reaction can be avoided if the cytidine residues are first deprotected with methonolic alkylamine at room temperature for 5 hours prior to ClϕU conversion and rA, rG deprotection at 55°C for 8 hours. The majority of protected cytidine is deprotected at room temperature resulting in very little mis-incorporation by nucleophilic alkylamine substitution into the undesired substituted cytosine.

During our initial attempts at convertible RNA synthesis, it was speculated that we consistently had varying amounts of base protected residues remaining after deprotection with cystamine disulfide. Visualization of the RNA oligonucleotides by UV shadowing showed the presence of smears in the gel, leading us to believe that the RNA was not completely deprotected. Therefore, we chose to use phenoxyacetyl (PAC) phosphoramidites in place of the standard β-cyanoethyl (CE) phosphoramidite base protecting groups. Iso-propyl-PAC rG, Acetyl rC, and PAC rA are more labile to base deprotection and therefore better suited for deprotection with an alkyl amine.

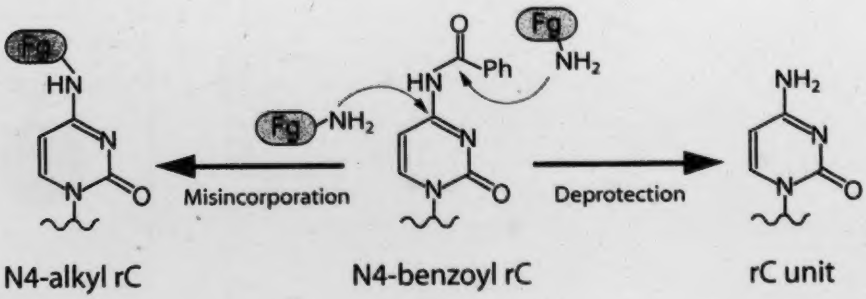


Figure II-2. Deprotection of benzoyl protected cytidine. Deprotection of N4-benzoyl rC with alkylamines can result in the deprotected nucleotide (right; rC) or mis-incorporation of the alkylamine from attack at the C4 position (left; N4-alkyl rC). The unwanted side reaction can be avoided by deprotection at room temperature.

We decided that the amine cystamine disulfide would be used for deprotection and conversion of modified RNAs. Specifically, the bifunctional amine cystamine disulfide was to be used as a precursor to derivatization of RNA oligonucleotides. Reduction of the disulfide bond leads to a sulfhydryl containing tether; the reactive thiol functionality can be derivatized with various electrophiles. After deprotection and conversion of the RNA oligonucleotides by cystamine disulfide, the solutions were neutralized with acetic acid in methanol, lyophilized to dryness and the salt crushed into a fine powder. The RNAs were then desilylated with t-butyl ammonium fluoride (TBAF) in tetrahydrofuran (THF) in the presence of the amine salt and desalted on a C18 Waters sep-pak cartridge to afford crude RNA. The desired RNAs were purified by denaturing PAGE to give excellent yields of fully deprotected functionalized RNA.

II-3. Materials and Methods.

II-3.1. Synthesis of 4-O-(4-chlorophenyl) uridine phosphoramidite.

All chemical reagents were obtained from Aldrich Chemical Company Inc., with exception of 2-cyano-N,N-diisopropylaminochlorophosphite from Sigma chemicals. Silica gel (0.03 - 0.07 mm) for flash chromatography was purchased from Acros Incorporated. Solvents were obtained from Caledon Laboratories Ltd., and were distilled as follows; THF from sodium/benzophenone; pyridine, dichloromethane, TEA from calcium hydride. TLC plates, pre-coated, 0.25 mm silica gel 60 with fluorescent indicator (40 × 80 mm) were obtained from Rose Scientific Limited. ^{1}H NMR performed on Varian Gemini 200 (200 MHz.) and ^{31}P NMR on Varian Gemini 300 (300 MHz.). All reactions were carried out under nitrogen atmosphere unless otherwise specified.

2',3',5'-tris-O-triethylsilyluridine (2). Uridine (5.0 g, 20.47 mmol) and imidazole (6.95 g, 102.4 mmol) were dissolved in 50 mL of dry dimethyl formamide (DMF). The resulting solution was cooled to 0°C and chlorotriethylsilane (13.75 mL, 81.9 mmol, 4.0 equiv.) was added drop-wise over ten minutes. The solution was then allowed to warm to room temperature for six hours. The solution was poured into 500 mL of diethylether and washed with 10% (w/v) aq. LiBr (4 × 150 mL) to remove DMF. The ethereal solution was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to a clear, colorless oil. Silica gel chromatography in 8:2

hexanes/EtOAc (R_f (**2**) = 0.82, 75:25 hexanes/EtOAc) yielded 11.562 g (19.70 mmol, 92 %) of **2** as a white foam.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.8 (s, 1H, H3-imino), 8.09 (d, 1H, H6), 5.85 (d, 1H, H1'), 5.67 (dd, 1H, H5), 4.10 (m, 2H, H2'/H3'), 4.05 (m, 1H, H4'), 3.95 (m, 1H, H5'/H5"), 3.75 (m, 1H, H5'/H5"), 0.93 - 0.99 (m, 27H, TES methyls), 0.58 - 0.68 (m, 18H, TES methylene).

2', 3', 5'-tris-O-triethylsilyl-4-O-(2,4,6-triisopropylbenzenesulfonyl)uridine (3). 2', 3', 5'-tris-O-triethylsilyluridine (**2**) (9.30 g, 15.84 mmol) and 4-dimethylaminopyridine (198 mg) were dissolved in 79.1 mL of dry, distilled CH_2Cl_2. The solution was chilled to 0°C, followed by the addition of triethylamine (13.5 mL). After thirty minutes, 2,4,6-triisopropylbenzenesulfonyl chloride (8.19 g, 27.08 mmol, 1.71 equiv.) was added. The reaction was stirred at 0 °C for one hour, then at room temperature for an additional eight hours. Over this time the reaction turned a deep red color, with the formation of a slushy precipitate. The crude mixture was poured into 1:1 hexanes/EtOAc and filtered. The filtrate was concentrated *in vacuo* to yield a brownish-yellow oil. Silica gel chromatography in 95:5 hexanes/diethylether (R_f (**3**) = 0.69, 9:1 hexanes/diethylether) yielded 10.00g (11.75 mmol, 74%) of **3** as a white solid.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.67 (d, 1H,H6), 7.20 (s, 2H, Trisyl aromatic), 6.00 (d, 1H, H5), 5.67 (s, 1H, H1'), 4,22 (septet, 2H, Trisyl 2,6 methines), 4.12 - 4.03 (m, 3H, H2'/H3'/H4'), 4.05 (m, 1H, H5'/H5"), 3.77 (m, 1H, H5'/H5"), 2.90 (heptet, 1H, Trisyl 4 methine), 1.30 (d, 6H, Trisyl 4 methyls), 1.26 (d, 12H, Trisyl 2,6 methyls), 0.99 - 0.91 (m, 27H, TES methyls), 0.70 - 0.58 (m, 18H, TES methylenes).

2',3',5'-tris-O-triethylsilyl-4-O-(4-chlorophenyl)uridine (4). 2', 3', 5'-tris-O-triethylsilyl-4-O-(2,4,6-triisopropylbenzenesulfonyl) uridine (**3**) (10.0 g, 11.75 mmol) and 4-chlorophenol (7.55 g, 58.7 mmol, 5.0 equiv.) were dissolved in 68.6 mL of dry distilled CH_2Cl_2 at 0°C. Anhydrous trimethylamine gas was bubbled through the solution at 0°C for fifteen minutes. While the solution was still cold, freshly distilled triethylamine (6.51 mL) was added. After 1.5 hours the solution was allowed to warm to room temperature. The flask was kept sealed to ensure the retention of the trimethylamine. Within three hours a white precipitate had formed. After a total of six hours, the excess trimethylamine was rotovaped off and the remaining solvent removed *in vacuo*. The reaction mixture was dissolved in 9:1 hexanes/diethylether to produce a thick viscous solution. Silica gel chromatography of the crude reaction mixture in hexanes was performed to remove the solid precipitate. Silica gel chromatography in 9:1 hexanes/diethylether (R_f (**4**) = 0.31, 9:1 hexanes/diethylether) yielded 7.3 g (10.48 mmol, 89%) of **4** as a yellow oil.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.66 (d, 1H, H6), 7.29 (d, 2H phenyl H3), 7.11 (d, 2H, phenyl H2), 6.05 (d, 1H, H5) 5.72 (s, 1H, H1'), 4.11 (m, 1H, H2'/H3'/H4'/H5'), 3.82 (m, 1H, H5"), 1.04 - 0.92 (m, 27H, TES methyls), 0.75 - 0.59 (m, 18H, TES methylenes).

4-O-(4-chlorophenyl)uridine (5). 2', 3', 5'-tris-O-triethylsilyl-4-O-(4-chlorophenyl) uridine (**4**) (7.3 g, 10.48 mmol) was dissolved in 78.9 mL of THF. Distilled deionized water (26.3 mL) and glacial acetic acid (26.3 mL) were added. The solution was allowed to stir at room temperature for thirteen hours until reaction appeared complete by TLC. The crude reaction solution was concentrated *in vacuo* to yield a clear, colorless oil. Silica gel chromatography with 11:1 CH_2Cl_2/methanol (R_f (**5**) = 0.44, 9:1 CH_2Cl_2/methanol) yielded 3.1g (8.74 mmol, 86%) of **5** as a white solid.

^{1}H NMR (DMSO-d_6, 200 MHz): δ (ppm) 8.51 (d, 1H, H6), 7.49 (d, 2H, phenyl H3), 7.22 (d, 2H, phenyl H2), 6.32 (d, 1H, H5), 5.72 (d, 1H, H1'), 3.98 (m, 2H, H2'/H3'), 3.91 (m, 1H, H4'), 3.75 (m, 1H, H5'/H5"), 3.61 (m, 1H, H5'/H5").

5'-O-(4,4'-dimethoxytrityl)-4-O-(4-chlorophenyl)uridine (6). 4-O-(4-chlorophenyl)uridine (**5**) (2.213g, 6.233 mmol) and 4-dimethylaminopyridine (85 mg) were dissolved in 38 mL of dry pyridine. The solution was cooled to 0°C, followed by the addition of 4,4'-dimethoxytrityl chloride (2.539 g, 7.49 mmol, 1.20 equiv.). The solution was stirred at room temperature for eight hours. After this time, the solution was cooled back to 0°C and quenched by the addition of 15 mL methanol. The solution was concentrated *in vacuo* and rotovapped down from toluene (100 mL) to yield a dark yellow oil. The oil was redissolved in 150 mL CH_2Cl_2 and washed with 5% (w/v) aq. $NaHCO_3$ (2 × 150 mL). The organic portion was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to yield a golden oil. Silica gel chromatography (silica was pretreated with 5% (v/v) triethylamine in 98:2 CH_2Cl_2/methanol) with 98:2 CH_2Cl_2/methanol (R_f (**6**) = 0.82, 9:1 CH_2Cl_2/methanol) yielded 2.703 g (4.11 mmol, 63.9%) of **6** as a pale brown foam.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.14 (d, 1H, H6), 7.33 (d, 2H, phenyl H3), 7.38 - 7.23 (m, 5H, DMT H2"/H3"/H4"), 7.26 (d, 4H, DMT H2/H2'), 7.07 (d, 2H, phenyl H2), 6.83 (d, 4H, DMT H3/H3'), 5.82 (d, 1H, H1'), 5.76 (d, 1H, H5), 4.38 (m, 1H, H2'), 4.28 - 4.25 (m, 2H, H3'/H4'), 3.79 (s, 6H, DMT methoxy), 3.52 - 3.43 (m, 2H, H5'/H5").

5'-O-(4,4'-dimethoxytrityl)-2'-O-tert-butyldimethylsilyl-4-O-(4-chlorophenyl)uridine (7). 5'-O-(4,4'-dimethoxytrityl)-4-O-(4-chlorophenyl)uridine (**6**) (2.0 g, 3.04 mmol) was dissolved in 11.7 mL of dry THF, to which had been added triethylamine (0.60 mL). After twenty minutes of stirring under nitrogen, *tert*-butyldiethylsilyl chloride (551 mg, 3.22 mmol, 1.06 equiv.) and silver nitrate (551 mg) were added to the solution. After stirring for four hours, the reaction mixture was filtered to remove silver salts. The filtrate was then diluted to 200 mL with ethyl acetate and washed with 5% (w/v) aq. $NaHCO_3$ (2 × 150 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to yield a pale foam. Silica gel chromatography (silica was pretreated with 5% (v/v) triethylamine in hexanes) with 75:25 hexanes/EtOAc (R_f (**7**) = 0.34, 75:25 hexanes/EtOAc) yielded 0.5063 g (0.638 mmol, 22%) of **7** as a white foam.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.51 (d, 1H, H6), 7.33 (d, 2H, phenyl H3), 7.31 (d, 4H, DMT H2/H2'), 7.43 - 7.26 (m, 5H, DMT H2"/H3"/H4"), 7.07 (d, 2H, phenyl H2), 6.85 (d, 4H, DMT H3/H3"), 5.81 (s, 1H, H1'), 5.64 (d, 1H, H5), 4.39 (m, 1H, H2'), 4.26 (m, 1H, H3'), 4.07 (m, 1H, H4'), 3.80 (s, 6H, DMT methoxy), 3.62 - 3.53 (m, 2H, H5'/H5"), 0.09 (TBS t-bu methyls), 0.30 (s, 3H, TBS methyl), 0.18 (s, 3H TBS methyl).

5'-O-(4,4'-dimethoxytrityl)-2'-O-tert-butyldimethylsilyl-4-O-(4-chlorophenyl)uridine-3'-O-(2-cyanoethoxy-N,N-diisopropylamino) phosphoramidite (8). 5'-O-(4,4'-dimethoxytrityl)-2'-O-*tert*-butyldimethylsilyl-4-O-(4-chlorophenyl) uridine (**7**) (0.5063 g, 0.638 mmol) was dissolved in 4.65 mL of dry THF under nitrogen atmosphere. Anhydrous N, N-diisopropylethylamine (0.679

mL) was added to the solution at room temperature, followed by addition of 2-cyanoethyl-N,N-diisopropylaminochlorophosphite (0.29 mL, 1.31 mmol, 2.06 equiv.). The solution was allowed to stir at room temperature. Within one hour, a white precipitate had formed. After six hours, the reaction mixture was poured into 100 mL of ethyl acetate and then washed with 5% (w/v) aq. $NaHCO_3$ (2 × 100 mL). The organic layer was dried over anhydrous sodium sulfate, filtered and concentrated *in vacuo* to a pale oil. Silica gel chromatography (silica was pretreated with 5% triethylamine in hexanes) in 6:4 hexanes/EtOAc (R_f of diastereomers of (**8**) = 0.81, 0.78, 1:1 hexanes/EtOAc). Concentrated *in vacuo* to yield 480 mg (0.483 mmol, 76%) of **8** as an off white foam.

^{1}H NMR ($CDCl_3$, 200 MHz): δ (ppm) 8.57 (d, H6, d1), 8.49 (d, H6, d2), 7.33 (d, 2H, phenyl H3), 7.29 (d, 4H, DMT H2/H2'), 7.46 - 7.24 (m, 5H, DMT H2"/H3"/H4"), 7.07 (d, 2H, phenyl H2), 6.84 (d, 4H, DMT H3/H3'), 5.82 (s, H1', d2), 5.73 (s, H1', d1), 5.61 (d, H5, d1), 5.50 (d, H5, d2), 4.36 - 4.31 (m, 3H, H2'/H3'/H4'), 3.80 (s, 6H, DMT methoxy), 3.72 (m, 2H, cyanoethyl H1/H1'), 3.53 (m, 2H, H5'H5"), 3.72 - 3.50 (m, 2H, diisopropylamino methines), 2.40 (t, 2H, cyanoethyl H2/H2'), 1.24 - 1.00 (m, 12H, diisopropylamino methyls), 0.89, 0.84 (s, 9H, TBS t-bu methyl), 0.25, 0.24, 0.13, 0.12 (TBS methyls).
^{31}P NMR δ (ppm) 152.5, 148.9.

II-3.2. Synthesis of functionalized RNA.

Synthesis of aminoethanethiodisulfide (cystamine disulfide). Cystamine dihydrochloride (25 g, 0.110 mol) was added to 200 mL 8 M KOH. The resulting solution was extracted with CH_2Cl_2 (4 × 250 mL). The extracts were dried over sodium sulphate, concentrated *in vacuo* to yield a pale golden-brown oil of aminoethanethiodisulfide (determined density: 1.168 g/mL). The stock solution was kept sealed and stored at -20°C.

Modified RNA oligonucleotide synthesis and deprotection. All RNA synthesis were carried out using Applied Biosystems 392 DNA/RNA synthesizer with a conductance trityl monitor. Synthesizer reagents were supplied by Perkin Elmer with the exception of ribonucleoside phosphoramidites and the nucleoside-controlled poured glass (CPG) solid supports which were obtained from CPG incorporated. PAC-phosphoramidites were obtained from Amersham Pharmacia Biotech. The RNAs were synthesized on 1.0 μmol RNA synthesis cycle using a 1.5 μmol scale of 500 Å CPG resin (45 mg DMT-rA(Bz)-2'-tBuSi-CPG, (32 μmol/g)). A standard coupling time of 15 minutes for all coupling reactions was employed. All phosphoramidites were dissolved to a concentration of 0.1 M in anhydrous acetonitrile. Synthesis of all oligonucleotides proceeded with average stepwise yields of greater than 95%.

Following automated DNA/RNA synthesis, resin bound unmodified oligonucleotides were suspended in 1.5 mL of 3:1 ammonium hydroxide in ethanol in screw cap Eppendorf tubes for deprotection. The deprotection was carried out in 55°C water bath overnight. The reaction was centrifuged to spin down solid support material and the supernatant transferred to a clean Eppendorf tube. The pellet was washed with 300 μL ethanol, vortexed and centrifuged again. The resulting supernatants were combined and concentrated under vacuum using a speed-vac concentrator. The tBDMS groups were deprotected by adding 300 μL of 1.0 M TBAF in THF. The pellet was fully dissolved and deprotected in the dark for 24 hours.

RNA sequences that contained 4-chlorophenyl uridine were deprotected with 0.5 mL of 4.0 M cystamine disulfide in anhydrous methanol. The freshly synthesized RNA was placed into screw cap Eppendorf tubes with base and deprotected at room temperature for 5 hours, followed by deprotection at 55°C for an additional 8 hours. The solutions were centrifuged to spin down the solid support material; the supernatant was removed and subsequently treated with 240 μL 17.4 M acetic acid in methanol to neutralize the amine used in deprotection. The neutralized oligonucleotide solution was speed-vac concentrated to a hard yellow salt that was crushed with a glass rod into a fine powder, and then treated with TBAF/THF for deprotection as described above.

Oligonucleotide purification and quantification. Deprotected crude oligonucleotides were desalted using a C-18 Analytical Sep-Pak column (Waters). The column was washed with 15 mL Acetonitrile, 30% (v/v) Acetonitrile-100 mM triethylamonium bicarbonate (TEAB) solution, 100 mM TEAB, followed by 25 mM TEAB. The crude RNA/TBAF reaction was quenched with 5 mL of 1 M TEAB and then brought up to 50 mL with deionized water and immediately passed through the column. The column was then washed with 20 mL of 25 mM TEAB and the RNA was eluted with 10 mL of 30% (v/v) Acetonitrile-100 mM TEAB solution and collected in eppendorf tubes. The fractions were frozen in dry ice, and then concentrated under vacuum using a speed-vac concentrator.

Desalted RNAs were purified by denaturing polyacrylamide gel electrophoresis. All gels were 20% (19:1 acrylamide/bisacrylamide) and run with 1×Tris-Borate-EDTA (TBE) running buffer. The crude RNA fractions were re-suspended in 200 μL distilled deionized water. Loading dye (0.4 mg/mL Bromo Phenol Blue/Xylene cyanole, 6.4 M urea, 0.4 M Tris-base, 0.4 M boric acid, and 100 mM Na_2EDTA) was added and the crude RNA solution, denatured on a 98°C hot plate for one minute, was electrophoresed at 35 mA for approximately five hours.

The product oligonucleotides were visualized by UV-shadowing, excised from the gel, crushed and soaked overnight at 37°C in 20 mL of 1 M TEAB with 1 mL phenol/chloroform. The product oligonucleotides were centrifuged to sink the polyacrylamide. The supernatant was applied to a C-18 Analytical Sep-Pak column (Waters), washed and eluted as described above. The fractions were frozen in dry ice and then concentrated under vacuum using a speed-vac concentrator. The RNA product fractions were pooled in 400 μL 0.3 M sodium acetate (pH 5.5), extracted with phenol/chloroform/isoamyl alcohol, chloroform/isoamyl alcohol, followed by ethanol precipitation. The RNA pellet was washed with 200 μL ethanol and dried in a speed-vac concentrator. The purified RNA was dissolved in distilled deionized water, quantified by UV absorbance at (A_{260}), and stored at –80°C.

II-4. References.

Allerson, C.R. (1996). Synthesis of a structurally constrained and functionalized tethered RNAs using convertible ribonucleosides. *Ph.D. Thesis Harvard*, 5-57.

Allerson, C.R., and Verdine,G.L. (1995). Synthesis and biochemical evaluation of RNA containing an intrahelical disulfide crosslink. *Chem Biol.* **10**, 667-75.

Cech, T., and Wang, J. (1992). Tertiary structure around the guanosine binding site of the *Tetrahymena* ribozyme. *Science* **256**, 526-529.

Eshaghpour, H., Soll, D., and Crothers, D. (1979). Specific chemical labeling of DNA fragments. *Nucleic Acids Res.* **7**, 1485-1495.

Dervan, P., and Yitzhak, T. (1993). Site specific enzymatic incorporation of an unnatural base, N-6-(6-Aminohexyl)-isoguanosine, into RNA. *J. Am. Chem. Soc.* **115**, 4461-4467.

Ghetu, A.F., Arthur, D.C., Kerppola, T.K., and Glover, J.N. (2002). Probing FinO-FinP RNA interactions by site-directed protein-RNA crosslinking and gelFRET. *RNA* **8**, 816-823.

Helene, C. (1989). Sequence specific recognition and cleavage of duplex DNA via triple helix formation by oligonucleotides covalently linked to a phenanthroline copper chelate. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9702-0706.

MacMillan, A.M., and Verdine, G. (1991). Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron* **47**, 2603-2616.

Rana, T., Tamilarasu, N., and Huq, I. (1999). Visualizing tertiary folding of RNA and RNA-protein interactions by a tethered iron chelate: analysis of HIV-1 Tat-TAR complex. *Nucleic Acids Res.* **27**, 1084-1093.

Qin, P.Z., Hideg, K., Feigon, J., and Hubbell, W.L. (2003). Monitoring RNA base structure and dynamics using site-directed spin labeling. *Biochemistry* **42**, 6772-6783.

Singh, K.K., Parwaresch, R., and Krupp, G. (1999). Rapid kinetic characterization of hammerhead ribozymes by real time monitoring of flourescence energy transfer. *RNA* **10**, 1348-1356.

Walter, N.G., Hampel, K.J., Brown, K.M., and Burke, J.M. (1998). Tertiary structure formation in the hairpin ribozyme monitored by flourescence energy transfer. *EMBO J.* **17**, 2378-2391.